Interplay of DNA Metabolic Pathways Involved in the Regulation of Gene Expression and Genomic Rearrangements

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Lia Kallenberger

von Amriswil TG

Promotionskommission

Prof. Dr. Josef Jiricny (Vorsitz und Leitung der Dissertation)

Prof. Dr. Ueli Grossniklaus

Prof. Dr. Primo Schär

Prof. Dr. Jörn Walter

Zürich 2017

Author	Lia Kallenberger Institute of Molecular Cancer Research Department of Biology Universität Zürich kallenberger@imcr.uzh.ch
Supervisor	Prof. Josef Jiricny Institute of Molecular Cancer Research Department of Biology Universität Zürich/ETH Zürich jiricny@imcr.uzh.ch
Committee Member	Prof. Ueli Grossniklaus Department of Plant and Microbial Biology University of Zurich grossnik@botinst.uzh.ch
Committee Member	Prof. Primo Schär Department of Biomedicine University of Basel primo.schaer@unibas.ch
Committee Member	Prof. Jörn Walter Zentrum für Human- und Molekularbiologie Universität des Saarlandes j.walter@mx.uni-saarland.de

Für meine Familie

SUMMARY

Tightly-controlled cell division is of critical importance for multicellular organisms to be able to grow and survive over long periods of time. The most demanding step in this process is the faithful replication of the genome. If a cell fails to copy its DNA molecule in an error-free manner, the genetic changes in the DNA can lead to cell death or, in the worst case, to uncontrolled cell division, resulting in genome instability and cancer.

The DNA is constantly challenged by exogenous and endogenous damaging agents, such as UV irradiation, water or reactive oxygen species that threaten the integrity of DNA. The resulting lesions are addressed by base excision repair (BER) and mismatch repair (MMR). The cell's repair pathways are, however, not only needed to correct damage in the DNA and mistakes arising during replication, but they are also essential for many physiological processes. BER and MMR were shown to play a role in antibody diversification during adaptive immune response and BER is implicated in DNA methylation dynamics to control gene expression.

In the first part of the thesis, we set out to investigate the role of BER and MMR in class switch recombination (CSR), a process that enables B-cells to produce antibodies with different effector functions. Using DNA substrates containing defined nicks and uracils, we mimicked the intermediates arising during CSR by the action of activation induced cytidine deaminase (AID). We found that the creation of DNA double strand breaks, a step that is essential for CSR, heavily depends on the BER enzyme uracil DNA glycosylase (UNG) and on the MMR protein MutSα.

In the second part of the thesis, I aimed to gain insights into the dynamic metabolism of DNA modifications that control gene expression by investigating the chromatin dynamics at the vitellogenin (*VTG*) gene regulatory region. The *VTG* gene represents a valuable model to study the events on chromatin leading to expression of the gene, as it is well-characterized and inducible by estradiol treatment. Demethylation events have been observed in the enhancer region in rooster DNA upon hormone treatment. In our cell line system, we saw no demethylation, but an increase in 5-hydroxymethylation, a product of oxidation mediated by the Ten-eleven Translocation proteins, implicated in DNA demethylation. In addition, the estrogen receptor α binding to its response element in the *VTG* enhancer was not altered by the presence of 5mC or 5hmC, but a shift of a methylation-sensitive factor was observed in nuclear extracts. *In vitro* methylation of a reporter plasmid under the control of the *VTG* regulatory region inhibited transcription, confirming an important role of methylation dynamics in the control of *VTG* expression. We found evidence that a CpG residing outside of the estrogen response element in an E-box in the VTG promoter is responsible for the methylation-sensitivity.

The data presented in this thesis illustrate the diverse functions of eukaryotic repair pathways. They show the great potential of these versatile mechanisms, but also point out the danger they pose to cells and multicellular organisms. If they are not processed correctly, the DNA modifications and rearrangements can lead to tumorigenesis, highlighting the importance of a deep understanding of the mechanisms involved.

ZUSAMMENFASSUNG

Streng kontrollierte Zellteilung ist sehr wichtig für multizelluläre Organismen, damit sie wachsen und über längere Zeit überleben können. Der anspruchsvollste Schritt dabei ist die zuverlässige Verdoppelung des Genoms. Wenn eine Zelle ihr DNS Molekül nicht fehlerfrei kopieren kann, können die resultierenden genetischen Veränderungen zum Zelltod oder noch schlimmer, zu unkontrollierter Zellteilung führen. Dies wiederum kann in Genominstabilität und Krebs resultieren.

DNS wird konstant von exogenen und endogenen schädlichen Wirkstoffen angegriffen, zum Beispiel UV-Strahlung, Wasser oder reaktive Sauerstoffverbindungen, die die Integrität der DNS bedrohen. Die resultierenden Läsionen werden von der Basen-Exzisions-Reparatur (BER) und der Fehlpaarungsreparatur (MMR) adressiert.

Die zellulären Reparaturmechanismen werden jedoch nicht nur gebraucht, um Schäden und Fehler in der Replikation zu korrigieren, sondern sind auch wichtig für viele physiologische Prozesse. BER und MMR spielen eine Rolle in der Diversifizierung von Antikörpern während der adaptiven Immunreaktion, und BER ist verwickelt in der Dynamik von DNS-Methylierungen, um Genexpression zu kontrollieren.

Im ersten Teil der These wollten wir die Rollen von BER und MMR in Klassenwechsel-Rekombination (CSR) untersuchen, einem Prozess, der es B-Zellen ermöglicht, Antikörper mit unterschiedlichen Effektor-Funktionen zu produzieren.

Der Einsatz von DNS-Substraten, die definierte Strangbrüche und Uracil-Basen enthalten - Läsionen, die von der aktivierungsabhängigen Cytidin-Deaminase (AID) erzeugt werden - ermöglichte es uns, die Zwischenprodukte von CSR nachzuahmen. Wir haben herausgefunden, dass die Erzeugung von DNS-Doppelstrangbrüchen, ein essentieller Schritt in CSR, vom BER-Enzym Uracil-DNS-Glykosylase (UNG) und vom MMR Protein MutSα abhängt.

Im zweiten Teil dieser Arbeit wollte ich einen Einblick in den dynamischen Metabolismus von DNS-Modifikationen in der Kontrolle von Genexpression gewinnen, indem ich die Chromatindynamik in der regulierenden Region des Vitellogenin (*VTG*)-Gens untersuchte. Das *VTG* Gen ist ein wertvolles Modellsystem, um die Ereignisse am Chromatin zu erforschen, die zur Expression des Gens führen, da das System gut charakterisiert ist und ausserdem mit Estradiol induzierbar ist. Demethylierung wurde in der Enhancer Region in der DNS von Hähnen nach einer Hormonbehandlung beobachtet. In unserer Zelllinie konnten wir keine Demethylierung sehen, aber einen Anstieg von Hydroxymethylierung, einem Produkt der Ten-eleven Translocation Proteine, die eine Rolle in der DNS-Demethylierung spielen. Zusätzlich haben wir herausgefunden, dass der Östrogenrezeptor seine Bindungsstelle im Enhancer des *VTG*-Gens unabhängig von Methylierung oder Hydroxymethylierung bindet, aber wir fanden einen methylierungsempfindlichen Faktor in nuklearen Extrakten. *In-vitro*-Methylierung eines Reporter-Plasmids unter der Kontrolle der *VTG*-regulativen Region führte zur Verhinderung der Transkription, was die Bedeutung von Methylierungsdynamiken in der Kontrolle der VTG Expression hervorhebt. Wir fanden Hinweise, dass ein CpG, ausserhalb des hormonreagierenden Elements in einer E-Box Sequenz des *VTG* Promotors, für die Methylierungs-Sensitivität verantwortlich ist.

Die in dieser These präsentierten Daten zeigen die verschiedenen Funktionen der eukaryotischen Reparaturwege. Sie zeigen das grosse Potenzial dieser vielseitigen Mechanismen, aber auch die Gefahr für die Zellen und multizellulären Organismen. Die DNS-Modifikationen und Umordnungen können zu Krebsentstehung führen, wenn sie nicht korrekt prozessiert werden, was die Bedeutung eines detaillierten Verständnisses der involvierten Mechanismen hervorhebt.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

8-oxo-dG	8-Oxo-2'-Deoxyguanosine
Alt-EJ	Alternative End Joining
AP-sites	Apurinic Sites
APE1	Apurinic/Apyrimidinic Endonuclease 1
APOBEC	Apolipoprotein B mRNA Editing Enzyme, Catalytic polypeptide
BER	Base Excision Repair
bHLH	basic Helix-Loop-Helix
bp	base pair
CE	Cytoplasmic Extracts
CSR	Class-Switch Recombination
dNTP	2'-Deoxyribonucleotide
DNA	Deoxyribonucleic Acid
DSB	Double Strand Break
EMSA	Electrophoretic Mobility Shift Assay
ER	Estrogen Receptor
ERE	Estrogen Response Element
GapDH	Glyceraldehyde-3-phosphate Dehydrogenase
H2AX	Histon H2AX variant
HIGM	Hyper-IgM Syndrome
HR	Homologous Recombination
LigI,III,IV	Ligase I, III, IV
MMR	Mismatch Repair
NE	Nuclear Extract
NHEJ	Non-Homologous End Joining
NR	Nuclear Receptor
nt	nucleotide
PAR	Poly ADP-Ribose
PARP1	Poly (ADP-Ribose) Polymerase 1
PARG	Poly (ADP-Ribose) Glycohydrolase
PCNA	Proliferating Cell Nuclear Antigen
PTM	Post-Translational Modifications
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RPA	Replication Protein A
RT	Room Temperature
SSA	Single-strand annealing
SSB	Single Strand Break
ssDNA	Single Strand DNA
SHM	Somatic Hypermutation
TET	Ten-Eleven Translocation
TF	Transcription Factor
TopI,II	Topoisomerase I, II

INTRODUCTION

GENOME STABILITY

All multicellular organisms face the challenging task of controlled cell division in order to be able to grow and survive over long periods of time. The most critical step in this process is the faithful replication of the genome. When a cell fails to copy its DNA molecule in an error-free manner, the genetic changes in the DNA sequence can lead, in the best case, to the controlled death of the cell, called apoptosis, but they can also lead to uncontrolled division of the cell, resulting in genome instability and cancer [1-3]. It is therefore of critical importance for the cell to ensure error-free replication of the genetic information and prevent any unintended changes in the DNA molecule. Cells possess several repair pathways that ensure the stability of the DNA sequence and its faithful duplication and propagation to the following generations, some of which will be discussed below.

Adding even more complexity to the metabolism of DNA, the cell's repair pathways are not only needed to correct damage in the DNA and mistakes arising during replication, but they are also essential for many physiological processes in the cell. It is of outmost importance for example that the DNA molecule can be dynamically modified for the regulation and control of gene expression. This occurs on an epigenetic level, meaning one level above the genetic code, but nevertheless involves transient modifications of the DNA. Even more striking, in addition to modifications of the DNA necessary for epigenetic regulation of gene expression, there are some examples where the DNA sequence itself is deliberately and irreversibly changed. In vertebrates for example, DSBs are introduced in a controlled manner into DNA during adaptive immune response, where extensive parts of the DNA sequence are excised [4, 5]. The role of DNA repair pathways in these physiological processes will be addressed extensively in this thesis.

Genome stability does therefore not imply that the DNA molecule is a rigid molecule that must not be changed. It rather describes the controlled and well-regulated dynamics of the DNA and proteins associated with it that enable a cell to divide many times without losing its characteristics.

In summary, the requirements of DNA metabolism are extremely high, ranging from dynamic and flexible DNA transactions to its long-term overall stability and error-free duplication.

DNA METABOLISM AND REPAIR

DNA is a remarkably stable molecule due in part to numerous inter-strand hydrogen bonds along the polymer. The polar amidino, guanidino and carbonyl groups of the complementary bases form a complex network consisting of hydrogen bonds with neighboring water molecules. Stability depends on the DNA sequence, as G/C base pairs are held together by three hydrogen bonds compared to only two in A/T pairs [6, 7]. But inter-strand hydrogen bonds are not the only factors contributing to the stability of DNA; stacking interactions between bases and base pairs were found to be the most important factors for the stability of the duplex [8]. They depend on the aromatic rings of the bases that form hydrophobic and electrostatic interactions. The degree of stability conferred by base

stacking also depends on the base sequence, as some combinations of base pairs form more stable interactions than others.

The high stability of DNA in a cell is somewhat unexpected, considering that, apart from exogenous sources like UV irradiation, there are also numerous endogenous agents that constantly attack DNA. The most important and most abundant ones are oxygen and water. Their contact with DNA is inevitable, but can lead to modifications that threaten the integrity of the DNA and the genome.

Molecular oxygen can react with unpaired electrons that escape the electron transfer chain, leading to the formation of reactive oxygen species (ROS). 1-2% of the electrons traveling through the respiratory chain are estimated to leave the chain that way. ROS are therefore generated frequently in metabolizing cells [9]. The most reactive species is the superoxide anion that is readily converted to hydrogen peroxide (H_2O_2). H_2O_2 is either decomposed to water or further converted to the strong oxidant hydroxyl radical [10, 11]. ROS impose an immense threat to DNA, because they react with the bases and the deoxyribosyl backbone causing mutagenic lesions, DNA breaks and DNA-protein cross-links [12, 13]. The formation of 7,8-dihydro-8-oxoguanine (8-oxo-dG) is particularly mutagenic in DNA, because its base pairing properties mimic those of a thymidine, causing the frequent misincorporation of adenine into the opposite strand [14, 15]. This leads to transversions from C/G to A/T, which are among the predominant somatic mutations in lung, breast and colorectal cancers [16]. It is therefore of outmost importance for the cell to remove 8-oxo-dG before replication.

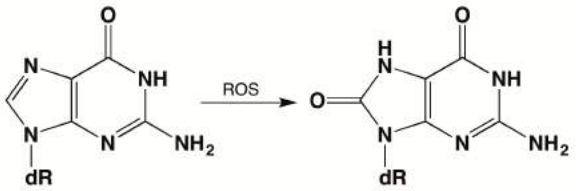


Figure 1: Oxidation of guanosine by ROS leading to the formation of 8-oxo-dG.

The other prevalent endogenous DNA damaging agent, water, also causes a variety of mutagenic DNA lesions. It hydrolyzes cytosine to uracil [17, 18] and 5-methylcytosine (mC) to thymine [19-21]. The latter reaction causes more complex problems, as the product is a *bona fide* DNA base and not an obvious damage product. In addition, adenine can be hydrolyzed to hypoxanthine, causing it to mispair with cytosine leading to A/T to G/C transitions [22, 23]. Water also causes protonation of guanine, eventually leading to its spontaneous loss through depurination [24] and, together with oxygen, it can generate thymine glycol and a plethora of other, minor, aberrant bases that need to be repaired by the cell to ensure the stability of DNA.

The third agent that spontaneously modifies DNA is S-adenosylmethionine (SAM); it acts as a gratuitous methyl group donor in a range of reactions, but it also gives rise to non-canonical modifications, such as 3-*N*-methyladenine, 7-*N*-methylguanine and 6-*O*-methylguanine [24, 25].

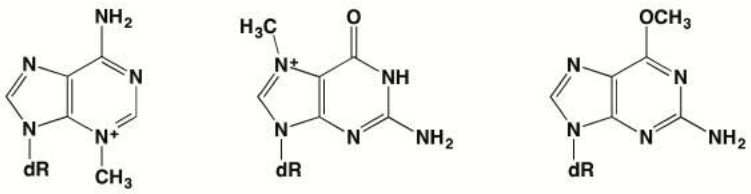


Figure 2: Alkylated purines generated in DNA by aberrant methylation. From left to right: 3-*N*-methyladenine, 7-*N*-methylguanine and 6-*O*-methylguanine.

Apart from arising through endogenous DNA damaging agents, cells also deliberately modify bases for regulatory purposes or to selectively modify the genome in a highly controlled manner. Some examples relevant for this work are mentioned briefly and will be discusses in detail later. SAM is not only a mutagenic agent in cells, but is also used to attach methyl moieties to the 5-position of cytosines, forming 5-methylcytosine (5mC) [26]. This so-called fifth base plays an essential role in many regulatory processes. In 2009, a family of oxidases, the Ten-eleven Translocation (TET) enzymes were discovered to oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) [27, 28]. The exact role of these additional cytosine species is under intense investigation.

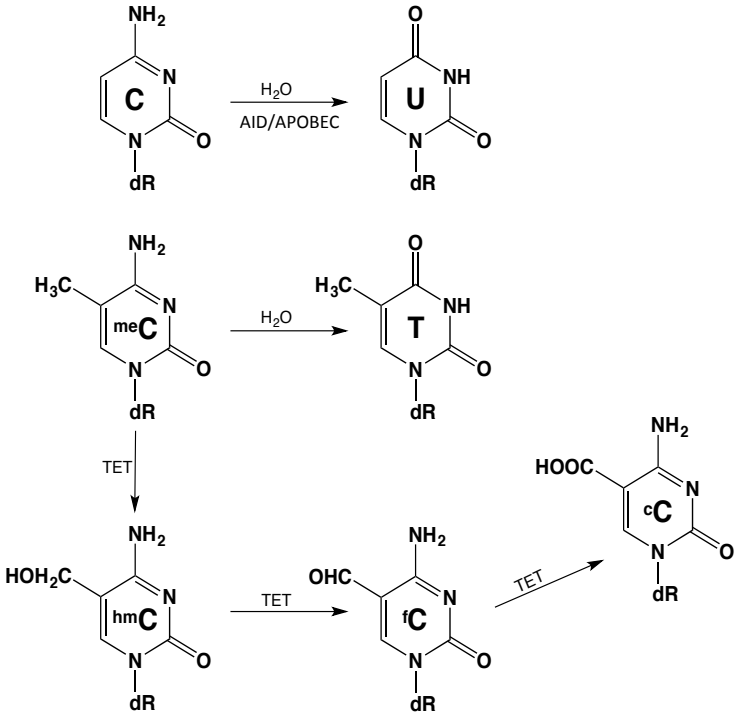


Figure 3: Spontaneous or enzymatic (by AID/APOBEC) hydrolytic deamination of cytosine (top row) gives rise to uracil. 5-Methylcytosine can be either deaminated directly to thymine (middle row), or enzymatically converted in a stepwise manner to hydroxymethylcytosine, formylcytosine and carboxycytosine by TET dioxygenases. When opposite G in double-stranded DNA, T, fC and caC can be excised by thymine DNA glycosylase.

Another enzyme family shown to deaminate cytosines in DNA consists of the APOBEC1/AID deaminases. Originally described to edit RNA, they were later shown to also modify DNA [29, 30]. Activation-induced cytidine deaminase (AID) is used by the cell to deaminate cytosines to uracils in the immunoglobulin locus in activated B-cells, to facilitate somatic hypermutation (SHM) and class-switch recombination (CSR).

Cellular repair pathways address all of the mentioned lesions, no matter whether they occurred spontaneously as damaging events or whether they were introduced by the cellular machinery. With the notable exception of 6-*O*-methylguanine, which is converted back to guanine by methylguanine methyltransferase (MGMT) by a process referred to as damage reversal [31], the above modifications are detoxified by base excision repair (BER), the resulting mismatches by mismatch repair (MMR), in the best case before the damage is converted to single strand breaks (SSBs) or, even worse, to double strand breaks (DSBs) or fixed in the genome as mutations. The repair pathways involved are discussed in more detail below.

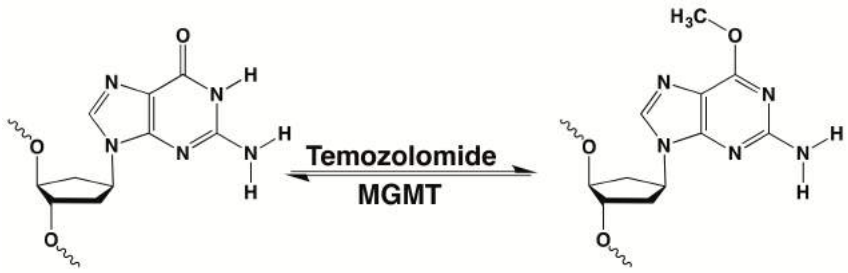


Figure 4: Damage reversal by MGMT, restoring guanine from 6-*O*-methylguanine generated for example by the alkylating chemotherapeutic Temozolomide.

Base Excision Repair

BER removes damaged bases arising from oxidation, alkylation and deamination and can also act on abasic sites resulting from non-enzymatic hydrolysis, as mentioned above. Damage on bases occurs very frequently, as it arises from spontaneous events caused by endogenous agents abundantly present in all cells. Most of these lesions cause severe problems when encountering replication, either because they cause mispairings that are turned into mutations during replication or because they block the replicative polymerases. A highly-efficient and sensitive process is therefore required to protect the genome from different kinds of DNA modifications. Damage-specific DNA glycosylases scan the DNA for lesions as they move along the double helix. More than ten different glycosylases are known in mammals, each recognizing a set of related lesions. The first one discovered was uracil DNA glycosylase (UNG) that efficiently excises uracils from single- and double-stranded DNA [32-34]. This lead to the discovery of many other damage-specific glycosylases that were grouped into four structurally distinct superfamilies; the uracil DNA glycosylases (UDGs), the helix-hairpin-helix glycosylases (HhH), the 3-methyl-purine glycosylases and the endonuclease VIII-like glycosylases [35]. Although the glycosylases differ in their structure and substrate specificities, they share a common mode of action. They recognize damaged bases by flipping them out of the helix into their respective active site pockets [36]. If fitting is successful, the damaged base is processed further. Despite being

highly-specialized enzymes, many of the glycosylases have overlapping functions leading to redundancies, as indicated by the mild phenotypes of single knockouts of the different glycosylases [37, 38]. The notable exception is thymine DNA glycosylase (TDG) that has been shown to be essential for embryonic development [39].

The minimal BER requires a monofunctional DNA glycosylase or a bifunctional DNA glycosylase/lyase, an AP endonuclease, a DNA polymerase and a DNA ligase [38, 40]. First, the damaged base is recognized and removed by the glycosylase, leaving behind an apurinic or apyrimidinic (AP) site, or, in the case of a bifunctional glycosylase, a 3' nicked AP site. The AP site is then processed by AP endonuclease 1 (APE1) that creates a single-stranded DNA nick 5' to the AP site or a single nucleotide gap when acting downstream from a glycosylase/lyase. Polymerase β then inserts the excised nucleotide. In the last step, the nick is sealed by XRCC1/ligaseIII [38]. The basic repair pathway is shown in figure 5, exemplified on UNG.

The excised nucleotide can be replaced either by short- or long-patch base excision repair. While short patch BER involves the replacement of a single nucleotide by polymerase β, followed by XRCC1/ligaseIII-dependent ligation of the nick, long patch BER is slightly more complex [41, 42]. It has been shown to additionally depend on proliferating cell nuclear antigen (PCNA) and the flap structure specific endonuclease 1 (FEN1) [43, 44], and involves the replacement of 2-6 nucleotides by polymerase β, ε or δ, via a strand displacement mechanism. The nicks in long-patch base excision repair are sealed by Ligase I. Due to the additional factors required, long-patch BER is mainly active in proliferating cells [45-47].

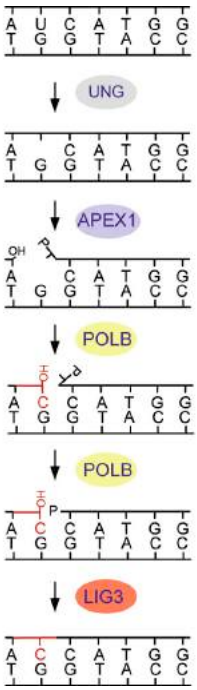


Figure 5: The basic base excision repair pathway exemplified on uracil DNA glycosylase. In a first step, the lesion, here a uracil, is recognized by the glycosylase. The base is excised, leaving behind an AP site that is addressed by an AP endonuclease, nicking the AP site 5'. Polymerase β inserts the missing nucleotide and removes the deoxyribose 5-phosphate using its lyase activity. Ligase III seals the nick. Taken from Robertson et al., 2009 [38].

Uracil DNA glycosylase superfamily

The UDGs are a big family of glycosylases consisting of six subfamilies, three of which are present in eukaryotes; the UNG subfamily, the mismatch-specific uracil DNA glycosylases (MUGs) and the single-strand-specific monofunctional uracil DNA glycosylases (SMUGs) [35, 48]. There are four mammalian uracil DNA glycosylases; UNG, SMUG1, methyl domain binding protein 4 (MDB4) and TDG that diverge considerably in their amino-acid sequence, but show similarity in the structure of their catalytic domains [49, 50]. They all excise deaminated Cs and incorporated Us from DNA [20].

Uracil DNA glycosylase

UNG efficiently excises uracils in double- as well as single-stranded DNA and acts independently of the base opposite the uracil [33]. UNG has an exceptionally high turnover rate of up to 1000 U excised per minute [51]. It possesses a very narrow catalytic cavity that provides high specificity towards uracils and the chemotherapeutic uracil analog 5-fluorouracil [52, 53]. Highly conserved amino-acid residues in the active pocket ensure this specificity. The N-glycosidic bond of the bound uracil is polarized by a conserved histidine and thus marked for nucleophillic attack. In addition, an aspartate deprotonates a water molecule that then attacks the deoxyribose, leading to the hydrolysis of the N-glycosidic bond [54]. Steric hindrance prevents purines from entering the active site pocket and the side chain of a tyrosine residue blocks the entry of thymine. Cytosine can enter the active site, but is not processed further due to unfavorable positioning [55].

There are two UNG isoforms in mammals, the mitochondrial form UNG1 and the nuclear form UNG2, arising from alternative splicing [56]. UNG1 is highly expressed in mitochondria-rich tissues like the heart or skeletal muscle [57]. Considering that most ROS arise during mitochondrial metabolism, it makes sense for the cell to have a specialized glycosylase in mitochondria. The nuclear form UNG2 has been more extensively studied. Whereas UNG1 is ubiquitously expressed, UNG2 expression is cell cycle regulated, the highest levels being observed in late G1/S. It has additionally been shown that in B-cells, UNG2 mRNA is only present after activation [58]. The differential expression levels can be explained by the usage of two different promoters for the transcription of the two isoforms [57]. Consistent with its high expression in S phase, UNG2 has been shown to bind to PCNA and partly co-localizes with replication foci [59]. Besides being the predominant glycosylase to repair U/A mispairs arising during replication, UNG2 is also the major repair enzyme addressing U/G mismatches arising through deamination of cytosines [33, 60]. Moreover, knockout studies have revealed a role of UNG2 in the adaptive immune response of B-cells.

UNG-deficient mice are viable and show only a slight increase in mutation frequency [61], but they are susceptible to developing B-cell lymphomas [62], hinting towards an important role of UNG in B-cell development. An altered mutational pattern in somatic hypermutation (SHM) was found. The ratio of nucleotide transition versus transversion at C/G sites was shifted significantly towards transition in the absence of UNG, whereas the ratio at A/T sites was unchanged. This is consistent with the fact that unrepaired uracils lead to transitions from C/G to T/A, whereas replication through UNG-generated abasic sites yields more transversions [63]. Rada *et al.* also observed an accumulation of IgM in *Ung*$^{-/-}$ mice, hinting towards a defect in CSR. Consequently, *in vitro* activated

B-cells deficient of UNG showed drastic reduction in CSR [64]. Humans with mutations in the UNG gene display the Hyper-IgM-syndrome (HIGM), characterized by increased serum IgM levels combined with low or absent serum IgG, IgA and IgE [58, 65]. HIGM is additionally linked to mutations in activation-induced cytidine deaminase [66], different cluster of differentiation (CD) genes [67] and mismatch repair proteins.

Both SHM and CSR are initiated by AID that locus-specifically deaminates cytosines to uracils. UNG and MMR proteins seem to be the main repair factors involved in the processing of these uracils. How the interplay of these two repair pathway leads to extensive mutagenesis and genomic rearrangements is still not well understood and is the focus of the first part of this thesis.

Thymine DNA glycosylase

The correction of G/T mismatches arising from deamination of 5mC in the CpG context has been debated [68-71]. Although the repair seems to be inefficient, depicted by a loss of CpG dinucleotides in evolution, it was clear that there had to be a way to faithfully repair G/T to G/C. Repair activity, directed towards a *bona fide* deoxyribonucleotide uncoupled from replication however, demands careful regulation. A repair activity on G/T mismatches has first been described in simian cells transfected with SV40 bearing mismatches. A bias of G/T repair was found, favoring correction towards G/C by 90% [72]. Further investigations with HeLa nuclear extracts showed a similar bias and provided evidence for a glycosylase-mediated repair, involving the formation of abasic sites [73, 74]. The enzyme responsible was subsequently purified from HeLa extracts and found to be thymine DNA glycosylase [75]. TDG also catalyzes the removal of uracil in G/U mismatches, supporting UNG in protecting the DNA from accumulating uracils [76, 77]. Other than UNG, TDG is mismatch-specific and therefore does not address A/U pairs arising from wrongly incorporated uracils, but only uracils in the G/U context, resulting from deaminated cytosines. Its mismatch specificity also implies that TDG only acts on dsDNA, in contrast to UNG that processes also ssDNA substrates. The enzymatic activities of UNG and TDG can be readily separated by the use of UGI, an inhibitory peptide specific for UNG [78]. Despite their overlapping substrate specificities, TDG was assigned into a new subfamily of UDGs, the MUGs.

Unlike UNG, MUGs have spacious, rather non-discriminating active site pocket, allowing a broader range of substrates. TDG is able to excise a wide variety of modified and damaged bases (Fig. 6), among many others 5-bromo- and 5-formyluracil, as well as hypoxanthine and thymine glycol [79-81]. TDG uses its N-terminus to make a non-specific contact with the DNA, scanning for non-Watson-Crick base pairing, therefore involving the opposite base in lesion recognition. Upon encountering a mismatch, the N-terminus clamp structure flips the base to be removed out of the double helix and stabilizes the G in the opposite strand [82-84]. An asparagine (Asn 140 in humans) acts as the catalytic residue, coordinating the nucleophilic attack on the N-glycosidic bond. TDG stays firmly bound to the resulting abasic site, translating in a low turnover rate [85]. TDG dissociates only upon modification with small ubiquitin-like modifier (SUMO) on its C-terminus, leading to a conformational change that relaxes the binding of the N-terminal clamp to DNA [86, 87].

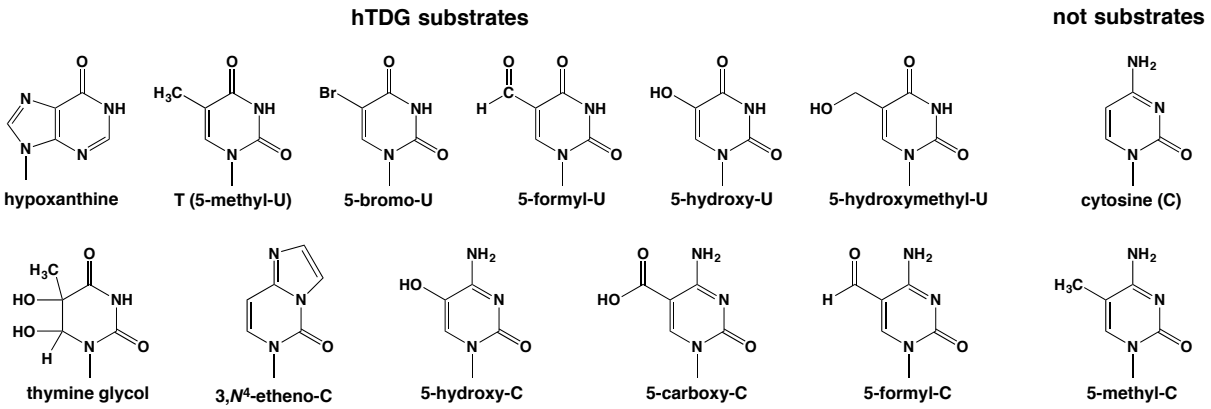


Figure 6: TDG substrate spectrum.

TDG displays an opposite regulation to UNG2 during cell cycle, with the highest expression in G1-phase and the lowest in S-phase [88]. Considering its activity towards thymine in G/T mismatches the importance of its absence during S-phase is obvious. It is not able to distinguish which nucleotide has been wrongly incorporated during replication in a G/T mismatch. It was found that TDG is specifically degraded during S-phase through a conserved PCNA-interacting peptide (PIP)-degron [89], first described in chromatin licensing and DNA replication factor 1 (CDT1). CDT1 and other degron-containing proteins are marked for degradation by the E3 ubiquitin ligase $CRL4^{Cdt2}$ upon interaction with PCNA. This depicts an efficient way to couple DNA synthesis and proteolysis, thereby preventing re-replication [90, 91] or in the case of TDG, erroneous repair of G/T mismatches arising from Gs, incorporated opposite Ts.

Aside from its important role in DNA repair, TDG has been repeatedly linked to gene regulation and DNA demethylation [92, 93], especially since the discovery of the TET proteins and their oxidation activity towards 5mC. TDG displayed glycosylase activity towards the TET products 5fC and 5caC [27, 94-97]. Interestingly, TDG is the only glycosylase with a lethal phenotype in KO mice that is dependent on its catalytic activity, but not linked to canonical base damage. This affirms an essential role of TDG aside from its known DNA repair activity [39, 98]. The involvement of TDG in transcriptional regulation and DNA demethylation will be discussed below.

Methyl-CpG binding domain protein 4

Although structurally distinct from the UDG family, the HhH gylcosylase methyl-CpG binding domain protein 4 (MBD4) is mentioned here briefly, due to its overlapping substrate specificities with UNG and TDG and to its role in methylation dynamics.

MBD4 was discovered in a database search for proteins with a methyl-CpG-binding domain (MBD) resembling MeCP1 (=MBD1) and 2, together with MBD2 and 3 [99]. All of the mentioned proteins bind methylated DNA. MBD1, MBD2 and MBD4 proteins specifically bind methylated DNA *in vitro*,

preferably double-stranded, symmetrically methylated DNA. Reduced binding of MBD4 to hemimethylated oligonucleotides was observed. The binding of MBD3 to methylated oligonucleotides was shown to be unspecific. MBD3 was found to lack a functional MBD, but it is an important component of the nucleosome remodeling and deacetylase (NuRD) repressor complex [100, 101]. MBD1 binds methylated DNA in cells, but it was shown to bind the same regions in DNMT KO cells, indicating nonspecific binding *in vivo*. The murine *Mbd* genes were expressed in all somatic tissues tested and all members, except MBD1, were present in embryonic stem cells (ESC).

MBD4 is the only known MBD protein that possesses another functional domain at the C-terminus, in addition to the N-terminal MBD. This C-terminal domain was described to be homologous to bacterial DNA repair enzymes, linking DNA repair and methylation [99]. Subsequently MBD4 was shown to efficiently excise thymine and uracil from mismatches, similar to TDG. Despite the similar substrate specificities, TDG and MBD4 show no amino-acid sequence homology, indicating that the proteins evolved from two different ancestors. MBD4 shows preference for the binding of 5mCpG/TpG mismatches. These mismatches are the main products of hydrolytic deamination of 5mCpG, hinting towards a role for MBD4 in minimizing mutations at 5mCpGs [102].

Consistent with the fact that MBD4 is endowed with 5mC binding- and glycosylase activities, there is evidence for a role of MBD4 in DNA demethylation in chicken [103] and zebrafish [104]. Interestingly, despite its obvious connection to DNA methylation, MBD4 is, unlike TDG and MeCP2, not embryonic lethal. Nevertheless, analysis of MBD4-deficient mice revealed elevated levels of C to T transitions at CpG sites translating into increased tumor susceptibility [105].

Mismatch Repair

MMR is a highly-conserved bidirectional process that repairs nucleotide incorporation errors and insertion-deletion loops (IDLs) occurring during DNA replication. MMR improves the fidelity of DNA replication by several orders of magnitude [106]. Cells deficient in MMR display greatly elevated mutation rates and microsatellite repeat instability, a hallmark of MMR deficient cancers (reviewed in [107]). Inherited defects in the MMR system underlie the disease hereditary nonpolyposis colon cancer (HNPCC), also called Lynch Syndrome, one of the most prevalent cancer syndromes in humans [108, 109].

In contrast to BER and other repair pathways, the substrate for MMR is undamaged DNA persisting only transiently in the DNA. In case the mismatches are not repaired before the next round of replication, half of the newly-synthesized strands will acquire a mutation that is no longer recognized as a lesion. It is therefore of great importance that MMR acts fast and that it is directed to the newly-synthesized strand to avoid mutagenesis. Apart from correcting nucleotide incorporation errors, MMR also has an important role in the protection against oxidative stress [110] and chemotherapeutics [111]. Moreover, it prevents aberrant recombination between non-identical sequences during meiotic recombination and in double-strand break repair (DSBR), as indicated by a hyper-recombinogenic phenotype of MMR-deficient cells [112, 113]. Operating as a lesion sensor, MMR proteins can also induce apoptosis and activate checkpoints of the cell cycle [114]. Surprisingly, considering its important function to ensure error-free replication, a non-canonical, mutagenic MMR

was implicated in antibody diversification during B cell development [115, 116] as well as in kidney and colon cancer cells upon encounter with DNA lesions [117, 118].

Canonical mismatch repair

Eukaryotic replicative polymerases display an error rate of about 10^{-4}-10^{-5}. The fidelity is enhanced about 100-fold by the 3'-exonuclease proofreading activity of the polymerases. MMR further increases replication fidelity by additional two to three orders of magnitude [106], resulting in a highly accurate process. IDLs, however, generated preferentially at repetitive microsatellite sequences, rely entirely on MMR, as they are not recognized by the proofreading activity of the polymerases. This translates into high microsatellite instability in the absence of MMR, the primary phenotype of MMR-deficient cells [119, 120].

MMR proteins are highly-conserved from bacteria to humans. They were first detected in *E. coli* [121], where the mechanism shows clear parallels to eukaryotic MMR, but also important differences. The first human MMR protein to be purified was MutSα. It was found in an attempt to identify enzymatic activity towards deaminated 5mC in DNA [122, 123]. Later, a minimal MMR system was assembled from purified human proteins [124].

In a first step in eukaryotic MMR, the mismatch or small IDL is recognized by MutSα, a heterodimer consisting of MSH6 and MSH2. The main substrates are 1-nucleotide IDLs and G/T mismatches [122]. MutSß, consisting of MSH3 and MSH2, recognizes larger IDLs. Upon mismatch binding, MutSα changes its conformation into a sliding-clamp, dependent on Mg^{2+}- and ATP [125]. MutSα then recruits MutLα, a heterodimer of MLH1 and PMS2 with endonucleolytic activity. This complex then moves along the DNA in search of a strand-discrimination signal. This step is essential for successful MMR and has long posed an enigma, as eukaryotes lack a MutH homologue, responsible for the strand discrimination in bacteria. To date, the signals directing MMR to the nascent strand are believed to be the 5' and 3' termini of Okazaki fragments and the 3' end of the leading strand [126-128]. There is recent evidence that nicks resulting from processed ribonucleotides, uracils and oxidized bases in the DNA act as supportive strand-discrimination signals [118, 129, 130]. These additional signals carry a threat however, as they may direct MMR to the wrong strand if they are generated in the DNA rather than being mis-incorporated during replication. One well-studied example is 6-*O*-methylguanine that arises from methylation of the DNA. Upon encounter with replication before being repaired by MGMT, this lesion either pairs with cytosine or thymine. Both base pairs are addressed by MMR that is targeted to the nascent strand harboring the unmodified base. This leads to 'futile' cycling of MMR-directed excision and resynthesis, causing cell cycle arrest and apoptosis ([131-134] and appendix).

As a next step in MMR, the endonuclease subunit PMS2 of the MutLα complex is activated by ATP and MutSα [135], further incising the nick-containing strand [136]. As a last step, exonuclease 1 (EXO1) is loaded and degrades the strand 5' to 3' past the mismatch. Consistent with the directionality of EXO1 it was shown that MutLα is only required when the strand-discrimination nick resides 3' of the mismatch, but is dispensable when the initial nick is 5' of the mismatch [118, 124]. The resulting single-stranded gap is filled-in by polymerase δ and the nick is sealed by ligase I.

Non-canonical mismatch repair

In contrast to the canonical, post-replicative MMR, non-canonical MMR acts independently of DNA replication and therefore does not discriminate between the two strands. The origin of substrates for MMR acting outside of S-phase are, similar to BER, damaged or modified bases that cause mismatched structures. However, non-canonical MMR does not specifically excise the damaged base, but excises a long stretch of DNA on the strand it has been directed to by a nick in close proximity. It is therefore difficult to imagine a scenario where MMR outside of S-phase would efficiently prevent mutations. Due to the fact that high-fidelity replicative polymerases able to fill the gap are scarce and dNTP pools suboptimal, it is in fact much more likely that MMR is mutagenic in this context.

Genetic studies implicated MMR, as well as monoubiquitylated PCNA (Ub-PCNA) and DNA polymerase eta (pol η), in the mutagenic processes somatic hypermutation (SHM) and class-switch recombination (CSR) during antibody diversification [137-143]. This non-canonical MMR, linked to PCNA monoubiquitylation was shown to be activated upon encounter with uracils or alkylation damage in B-cell extracts as well as in human embryonic kidney cells [117]. Ub-PCNA led to the recruitment of pol η, as shown in human cells [144].

SHM and CSR are both initiated by activation-induced cytidine deaminase (AID), which converts cytidines in immunoglobulin loci to uracils [145, 146]. Those uracil-guanine mismatches are substrates for BER and MMR and both seem to play a role in antibody diversification, as deficiencies of either process lead to impaired SHM and CSR [147]. The involvement of an error-prone translesion polymerase in combination with MMR acting randomly on both DNA strands represents a nice system for introduction of somatic mutations [148-150]. Consistent with this it was found that mutations at A:T sites depend on MMR [138] and on pol η [151]. However, the interplay of BER and MMR in CSR remains largely enigmatic, as mentioned above. It was hypothesized earlier that DSBs could be created by the collision of MMR degradation tracts with single-strand breaks (SSBs) created by BER at uracil sites [152]. This hypothesis will be discussed in detail later.

Single Strand Break Repair

Spontaneous base loss (primarily depurination) or incomplete BER can give rise to single-strand breaks (SSBs). SSBs occur relatively frequently, but are usually repaired by a rapid and efficient process. If left unrepaired, they are recombinogenic and can lead to the formation of double strand breaks (DBSs). The basic single-strand break repair (SSBR) pathway can be defined by four steps; SSB detection, DNA end processing, DNA gap filling and DNA ligation [153].

Detection of the lesion is carried out primarily by poly (ADP-ribose) polymerase-1 (PARP1). Upon binding to DNA, PARP1 activity is stimulated and catalyzes the generation of poly (ADP) ribose chains on itself and other acceptor proteins in close proximity. This first recognition and activation step is fast and PARP1 is recycled to its de-ribosylated state by poly (ADP-ribose) glycohydrolase (PARG) [154, 155]. The PARylation of accessory proteins then aids in recruiting further factors, like the scaffold protein XRCC1, needed to complete the repair. Depending on its origin the SSB may have

different termini that require different accessory repair proteins to restore conventional 3'-hydroxyl (3'-OH) and 5'-phosphate (5'-P) moieties needed for gap filling and ligation [153]. Most ends of SSBs can be repaired by polynucleotide kinase/phosphatase (PNKP) and APE1, whereas termini arising from incomplete BER typically need the lyase activity of polymerase β and flap endonuclease-1 (FEN1) in case of a modified 5'-end that the polymerase cannot remove [156]. Another type of lesion addressed by SSBR is the covalent attachment of topoisomerase 1 (Top1) to DNA. These lesions require tyrosyl DNA phosphodiesterase 1 (TDP1) to remove the adduct before gap filling and ligation can take place [157].

Once the termini of the SSB have been repaired, gap filling is carried out mainly by pol β. Similarly to BER, this occurs by introducing a single nucleotide (short-patch) or multiple nucleotides (long-patch), depending on the way the ends were processed. Ligase IIIα or ligase I seal the remaining nick in the backbone.

Double Strand Break Repair

In case the above-mentioned repair pathways fail to complete repair, the SSBs will be converted into DSBs when they encounter replication. DSBs are among the most dangerous lesions occurring in cells. When unrepaired, they can lead to deletions, translocations and chromosomal fragmentation, likely resulting in cell death or tumor formation [158]. About 10 DSBs occur per day in every dividing human cell [159], highlighting the need for faithful repair mechanisms. Eukaryotic cells evolved two major pathways to deal with these dangerous lesions; homologous recombination (HR) and non-homologous end-joining (NHEJ). In the absence of factors needed for the canonical pathways NHEJ and HR, cells can choose alternative repair mechanisms. Single-strand annealing (SSA) depends on RAD52 and involves annealing at long homologies, resulting in extensive deletions. Alternative end-joining (Alt-EJ), also called microhomology-mediated end joining (MMEJ), results in either deletions or insertions of short microhomologies (FIg. 7).

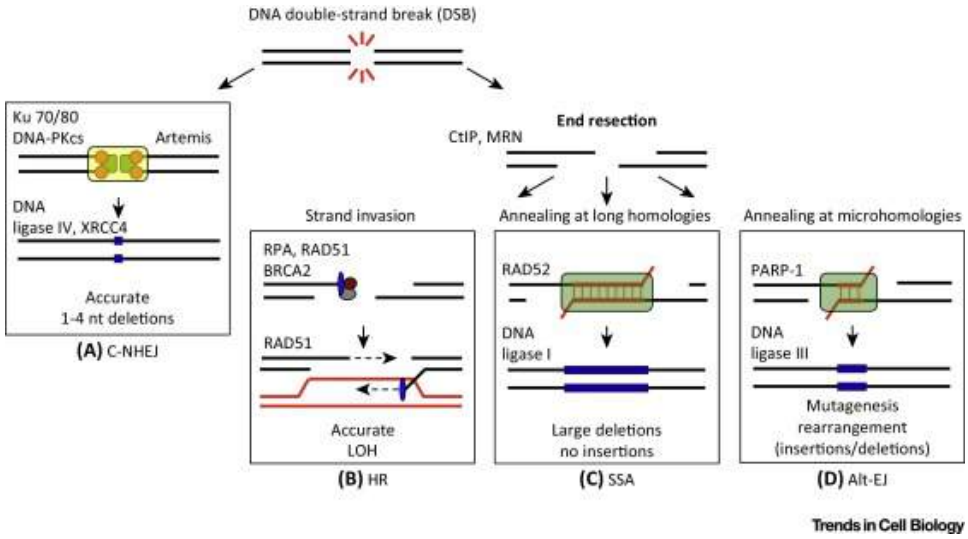


Figure 7 DNA double strand break repair pathways. Depending on the nature of the break and the cell cycle phase, the cells repairs DNA DSBs via canonical NHEJ (A), or it resects the DNA at the breaks, channeling the repair into HR (B), SSA (C) or Alt-EJ (D). From [160].

Homologous recombination

HR is essential for highly-accurate repair of DSBs. In addition, it has a role in inter-strand cross-link repair, supports telomere maintenance and is involved in the formation and resolution of cross-overs and chromosome segregation in meiosis.

As the name implies, HR relies on the presence of an intact homologous DNA molecule it can use as a template. This ensures a faithful and error-free correction of the lesion, but it also largely restricts this repair pathway to the S and G2 phases, where a homologous sister chromatid is present [161, 162]. Most HR genes were identified by displaying hypersensitivity towards DNA-damaging agents, such as ionizing radiation, when mutated [163, 164].

Upon DSB detection by the MRE11/RAD50/NBS1 (MRN) complex, histone H2AX is phosphorylated by ataxia telangiectasia mutated (ATM), or ataxia telangiectasia and Rad3-related (ATR) in response to replication stress, generating histone γ-H2AX. This phosphorylation serves as a platform to recruit other repair proteins [165]. It is followed by a 5' to 3' end-resection of the broken DNA ends, yielding 3' single-stranded overhangs. This resection step is critical to channel repair towards HR, as opposed to NHEJ that is carried out without extensive end resection. Resected DNA ends are poor substrates for the Ku proteins, the initiating enzymes in NHEJ. Cells are therefore restricted to continue along the HR- or other homology-mediated repair pathways, such as SSA or Alt-EJ [166].

The 3' single-stranded tail is first coated by replication protein A (RPA), which has been shown to have a positive as well as an inhibitory effect on HR. It helps to stabilize ssDNA generated during resection [167], but then suppresses the ATPase and recombinase activities of RAD51 if not removed [168, 169]. Therefore, RPA is replaced by RAD51 with the help of Breast Cancer type 2 susceptibility protein (BRCA2) in mammals, which interacts physically with RAD51 [170, 171]. RAD51 forms a nucleoprotein filament, consisting of about six molecules per 18-19 bases [172-174]. This presynaptic

filament captures a duplex DNA molecule and searches for homology, followed by RAD51-dependent strand invasion to form a D-loop intermediate. RAD51 assembly and homology search constitute the rate limiting steps of HR [174]. DNA polymerases extend the 3' end of the invading strand, followed by disengagement of the two strands. This process, called synthesis-dependent strand annealing (SDSA), results in non-crossover products [175]. Alternatively, double Holliday junctions can be formed that are resolved to form either non-crossovers or crossovers.

Non-Homologous End Joining

In contrast to HR, NHEJ has no need for a homologous template. It restores the integrity of the DNA double-helix by directly re-joining the two ends. Due to its applicability throughout the cell cycle, NHEJ is the more common pathway of choice in higher eukaryotes, even though it is error-prone and can result in genomic translocations. In fact, most of the genomic rearrangements occurring in our genome and eventually causing cancer are induced by NHEJ [176]. Apart from fixing pathogenic DSBs, NHEJ is important for the processing of physiologically-induced DSB, for example during antibody diversification [177]. Patients lacking NHEJ do not only show sensitivity to ionizing radiation, but they also suffer from a human severe combined immunodeficiency (SCID) [178, 179]. SCID is an autosomal-recessive disease, characterized by absence of circulating T and B lymphocytes, indicating a defect in V(D)J recombination [178, 179]. About 15% of SCID cases originate from defects in NHEJ, mostly in Artemis [180, 181]. CSR is also impaired in NHEJ-deficient cells, usually masked by the defect in V(D)J recombination [4].

NHEJ proteins show great adaptability, essential for the repair of a wide range of DNA end structures. In the first step, the proteins Ku70 and Ku80, competing with MRN for DSBs, form a heterodimer and bind to free, unresected DNA ends, marked by γ-H2AX [182]. The ends are thusly protected from degradation and the Ku proteins act as a recruiting platform for Artemis and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) [183]. Artemis has a nuclease activity and its recruitment leads to resection of the different DNA end configurations. The DNA-PKcs-Artemis complex has 5' -and 3-endonuclease, as well as hairpin opening and 5' exonuclease activity [181, 184]. The short-range resection of the ends is followed by binding of DNA polymerases μ or λ that repair the ends. The polymerases may generate small deletions and insertions, introducing further inaccuracy. Subsequently, the XLF/XRCC4/DNA ligase IV complex is recruited, which is able to ligate also incompatible ends and gaps, endowing the process with high flexibility, but also inaccuracy. XRCC4/ligase IV has in fact been shown to be the only complex needed for repair of DSBs in case the ends share perfect homology of at least 4 nucleotides [185, 186]. In case the classical NHEJ and HR fail, the cell can switch to SSA that relies on long homologies or an alternative end-joining pathway that relies on microhomologies.

Alternative- and Microhomology-Mediated End Joining

Alternative end-joining is a term that emerged recently to summarize the various sub-pathways that deviate in some aspects from the classical NHEJ pathway. Alt-EJ includes ligase IV-independent NHEJ,

Ku-independent NHEJ or DNA-PKcs-independent NHEJ [184]. It is also frequently called microhomology-mediated end joining (MMEJ), as short homologous sequences are used to tether the ends together in most of these sub-pathways. It is discriminated from SSA, which involves longer resection tracts and the annealing at long homologies. MMEJ was first described in Ku80-deficient yeast [187] and hamster cells [188]. Two studies in budding yeast provided further evidence for a Ku-independent repair mechanism [189, 190]. In addition, they revealed a role for MRE11, most likely by nucleolytic processing of the ends. Involvement of Pol η and ζ in the pathway was also postulated. In fission yeast, the participation of EXO1 in nucleolytic processing was shown and the MMR proteins PMS1 and MSH2 were attributed a role in resolving mismatches within the overlapping homologous tracts, therefore counteracting imperfect microhomologies [191].

For ligation, the MMEJ mechanism seems to rely more on DNA ligase I or the XRCC1/ligase III complex, in contrast to XRCC4/ligase IV acting in NHEJ [192]. Altered patterns of CSR bearing more microhomologies were observed in human patients with DNA ligase IV mutations [193]. A role for poly(ADP-ribose) polymerase-1 (PARP-1) was moreover postulated in MMEJ [194, 195], where it is believed to compete with Ku for the binding of free DNA ends. Indeed, it has be shown that PARP-1 binds DSBs with similar or even higher affinity than SSBs [196, 197].

ANTIBODY MATURATION AND DIVERSIFICATION

Our immune system is steadily challenged by a variety of constantly-evolving pathogens. The vast majority of intruders is detained by the innate immune system. This system is however not specific and cannot ensure long-lasting protective immunity to the host, like the adaptive immune response does. The production of newly-recombined immunoglobulins (Igs) is therefore necessary for an efficient host defense.

The main players in the adaptive immune system are T- and B-lymphocytes that recognize the infecting pathogens in a highly specific manner. They accomplish their task with the help of membrane receptors in the case of T-cells (cellular response) or with soluble antibodies called immunoglobulins (Igs) in the case of B-cells (humoral response). This work is focused on B-cells, as they are the only cells making use of SHM and CSR to increase the diversity of their antibodies.

Each immunoglobulin produced by B-cells comprises two light and two heavy chains, all four of which have a variable and a constant region (Fig. 8). The variable regions, containing the antigen-binding sites, are generated by recombination of the variable (V)-, diversity (D)- and the joining (J)-region (V(D)J recombination) in progenitor lymphocytes of the bone marrow. The two types of light chains (λ and κ) harbor several different variable regions in their genomic locus, 4, respectively 5, joining regions and no diversity regions, whereas the heavy chains are assembled from a choice of many variable regions, 6 joining regions and 25 diversity regions. In the case of the heavy chain, a D_H region is first joined to a J_H, before a V_H is juxtaposed to a D_HJ_H segment. The recombination events are initiated by the recombination activating genes 1 and 2 (RAG1 and RAG2), expressed exclusively in early B- and T-progenitors. They introduce DSBs in conserved non-coding DNA segments located adjacent to the coding sequence of the V, D and J segments. The DNA breaks are rejoined in a highly

orchestrated manner, involving NHEJ, as mentioned above. (Reviewed in [198-200]).

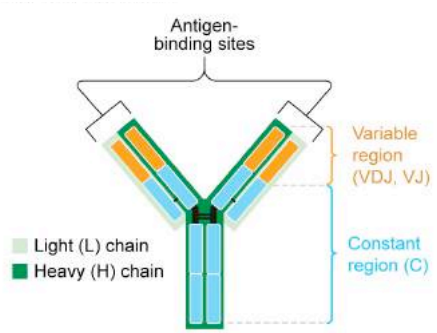


Figure 8: Structure of an antibody with two light and two heavy chains, each comprising a variable and a constant region. From [201].

The choices of V, D and J segments and combinations of different light and heavy chains, together with small insertions and deletions occurring during the NHEJ-mediated repair of the DSBs, creates the first layer of diversity. The antibodies expressed at this stage do not yet have very high affinity towards antigens, but it is high enough for the immune system to flexibly interact with many different antigens. To neutralize the antigens, however, very high affinity interactions are required [202]. This is achieved by SHM and CSR, processes that follow V(D)J recombination in B-cells, once they leave the bone marrow and migrate to peripheral lymphoid organs, where they form germinal centers ([5], FIg. 9). Both of these processes are initiated by AID upon encounter with antigens. Different mechanisms of resolving the AID-mediated lesions lead to very distinct outcomes.

SHM is the process by which B-cells introduce point mutations into the variable DNA segments of the immunoglobulin genes to produce higher-affinity Igs [203]. Mutations are generated at a rate about 10^6 greater than the spontaneous mutation rate in the genome [204]. Apart from AID, this highly-mutagenic process depends on BER and MMR, as discussed above.

CSR enhances the efficiency of the humoral response by changing the Ig isotype from IgM (or IgD) to the production of IgA, IgE and IgG, antibodies with different biological properties, but with the same antigen specificity [205]. The isotype of an Ig is defined by its constant region and defines the tissue distributions and effector functions of the antibody. The stimulus provoked by the antigen dictates the type of antibody the cell switches to. Like V(D)J recombination, CSR involves the deliberate formation of DSBs. The mechanism in CSR is less precise, but accordingly more flexible. It involves BER and MMR, similarly to SHM, but with a different outcome. Antibody diversification is an excellent system to study DNA modification upon different stimuli, influenced by sequence contexts, different enzymatic activities and interplay between repair pathways.

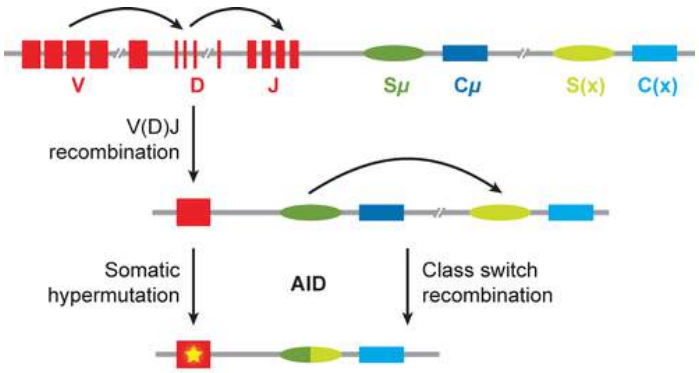


Figure 9 Mechanisms of immunoglobulin diversification schematically represented on the mouse IgH locus. One segment of each variable (V), diversity (D) and joining (J) are combined in a process called V(D)J recombination. Subsequently, upon activation of the B-cell, AID initiates somatic hypermutation of the newly recombined variable region and class switch recombination of the heavy chain constant region. From [206].

Activation-induced Cytidine Deaminase

AID plays an essential role in adaptive immunity by initiating both SHM and CSR. It is a member of the AID/APOBEC family of Zn-dependent cytosine deaminase-related enzymes, most of which are able to deaminate cytosine to uracil in RNA and in ssDNA. It was discovered in a cDNA screen comparing uninduced and induced lymphoma cells [146, 207].

AID was found to act solely on ssDNA [208, 209], a fact that has posed a challenge due to its role in targeting genomic DNA and has lead to extensive research. It was found that AID is directed to the target sequences by transcription, which generates transient ssDNA in the so-called transcription bubble [210, 211]. A direct interaction with replication protein A (RPA) has been found, possibly assisting in directing AID to its target sequences [212, 213]. AID deaminates cytosines on both ssDNA strands with a preference for WR**C** (W, A or T; R, purine; C, cytosine) sequence motifs [214, 215]. Its cytidine deaminase motif consists of two cysteines and a histidine residue for zinc coordination and a glutamic acid providing the proton for deamination [206, 216]. AID has very high affinity for ssDNA, resulting in low turnover rates. Another consequence of tight binding to DNA is the long time period AID resides bound to DNA. This enhances its role as a recruiting platform, giving accessory proteins time to process the deaminated cytosines [214, 217].

Interestingly, mutational studies functionally uncoupled the roles of AID in SHM and CSR. It was shown that while the catalytic function is essential for both processes, the C-terminus is additionally required for CSR [218, 219], whereas the N-terminus is necessary for SHM [220]. These findings clearly indicate that the steps following cytosine deamination are critical for the outcome of the process and are controlled, at least partially, by AID.

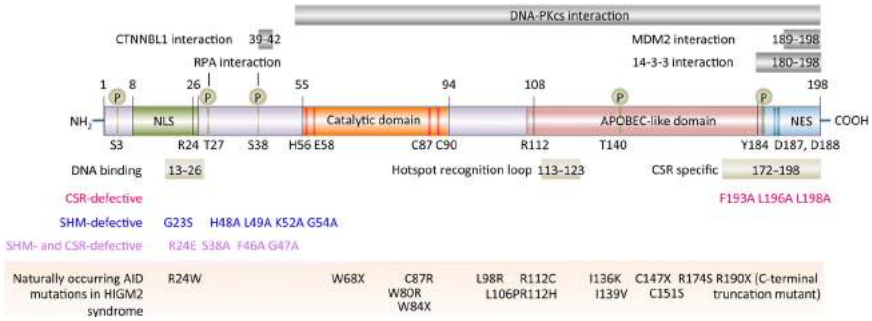


Figure 10: AID functional domains showing the naturally occurring mutations in the AID gene that are responsible for the autosomal recessive disorder HIGM. These mutations, depending on the domain they affect, cause defects in CSR and/or SHM. From [221].

Loss of function mutations in *AID* abrogate SHM and CSR, leading to the HIGM syndrome [66, 146, 222]. Without AID, the cell is not able to change the isotype from IgM to another class of immunoglobulin (Ig), resulting in an accumulation of IgMs in those patients. Together with the non-functional SHM, this severely impairs the humoral immune response, leading to persistent infections.

On the other hand, AID activity can have severe pathological side effects. The processes initiated by AID can lead to potentially transforming lesions such as genomic translocations provoked by the DSBs formed as CSR intermediates. Translocations involving the *Ig* locus, such as the well characterized *IgH-cMyc* translocation, were found to be a hallmark of Burkitt's lymphoma [223]. But also non-*Ig* genes can be targets of aberrant AID activity [224]. Elevated levels of AID were linked to autoimmune diseases, for example [225]. Interestingly, ubiquitous transgenic overexpression of AID has more severe effects on T-cells and other tissues than on B-cells, suggesting that B-cells have evolved protective mechanisms [226].

The sensitive balance of AID levels in B-cells has likely evolved as a trade-off between an effective and highly adaptive immune response and the risk of cancer or autoimmune diseases [202]. The regulation of AID is therefore essential and takes place on various levels. Apart from its transcriptional regulation, there is complex post-transcriptional regulation. The AID mRNA is bound by two miRNAs, miR-155 [223] and miR-181b [227], negatively regulating AID levels. To further limit its activity on DNA, AID is kept predominantly cytoplasmic (>90%), where it is very stable. After being transported into the nucleus [228-230], AID is constantly targeted to the proteasome. In addition, AID is regulated through phosphorylation on five residues that affects its localization and catalytic activity [231-234].

AID is a B-cell specific enzyme and therefore only B-cells and not T-cells are able to undergo SHM and CSR. Nevertheless, low levels of AID have been reported in the ovary, heart, prostate, lung and primordial germ cells, where the function of AID remains unclear [202].

Apart from converting cytosine to uracil, AID is also able to deaminate 5-methylcytosine to thymine, although with much lower efficiency [209, 235, 236]. This finding led to the very controversial discussion about the involvement of AID in DNA demethylation and epigenetic reprogramming [237, 238] that will be discussed later.

Class-Switch Recombination

In order to switch antibody isoform and guarantee an efficient immune response in different tissues, CSR of the constant heavy chain (C_H) region takes place. Each isotype possesses one, or in the case of IgG four, constant region; Cμ for IgM, Cδ for IgD, different Cγ for IgG3, 1, 2a and 2b, Cε for IgE and Cα for IgA. IgD is produced through alternative splicing of the IgM transcript. Before CSR takes place, IgM and IgD isoforms are produced. Five IgM molecules interact to form pentamers that can activate the complement system, but have low affinity for antigens. They are secreted, but cannot leave the vessels due to their large size. The switch from IgM/IgD to the high affinity antibodies is enabled by the joining of the chosen downstream C_H region to the previously-assembled VDJ region, and the excision of the intervening DNA fragment (Fig 11).

The different C_H regions are cytokine-inducible transcription units each consisting of a promoter, I_H-exons, a switch (S) region important for the mechanism of switching and C_H-exons. Specific cytokine-activated transcription factors (TFs) initiate the transcription through the Sμ (IgM) and the respective downstream S region, producing non-coding germline transcripts. S regions are 1-10 kb long, repetitive sequences 5' of every C_H region, except for Cδ (IgD) [205]. S regions display a high density of the AID targeting sequence WRC. Additionally, they consist of tandem repeats of 20-80 bp G-rich sequences, resulting in the generation of ssDNA R-loop structures that are more easily attacked by AID [49, 239]. This contributes to an accumulation of uracils in the S regions, whose processing results in DSBs. The DNA fragment between two breaks is spliced out as an extrachromosomal circle (Fig. 11). The exposed C_H terminus is then joined to the VDJ region, preferentially by NHEJ, but if need be also by the poorly-characterized alternative-NHEJ pathway that depends on microhomologies [240-242].

After successful class switching, the mono- or dimeric IgE, IgG and IgA are distributed systemically to tissues, where they mediate different biological effector functions [243]. The effectiveness of the humoral response is greatly increased after CSR, even though the antigen-binding sites of the antibodies remain unaltered.

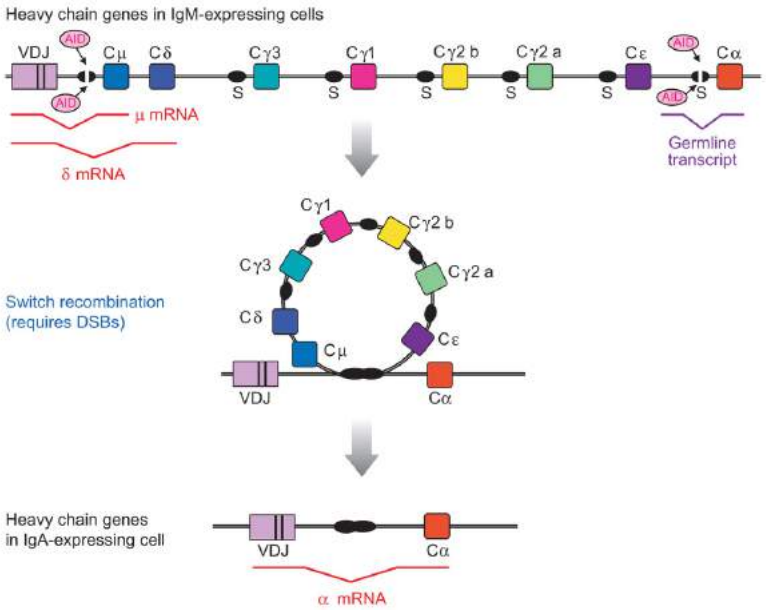


Figure 11: CSR from IgM- to IgA-expression. A promoter upstream of the VDJ region initiates transcription of the IgH locus yielding IgM (or IgD if alternativley spliced) antibodies before CSR had occurred. To enable CSR in activated B-cells, two DSBs have to be introduced with the help of AID; one in the very first S region (Cμ) and one in the specific S region the cell switches to (here Cα). The DNA in between is spliced out as a circular DNA fragment and the new immunoglobulin isotype can be transcribed. From [49].

The factors and pathways involved in CSR are well defined, being AID [66], BER with UNG [64, 244], MMR with MutSα [137, 139, 140, 142], and to a lower extent MutLα [245], and NHEJ factors, although the latter are not essential [4, 246]. Still, the exact mechanism of DSB induction and the specificity towards S regions in B-cells remain enigmatic. The current model postulates that MMR helps create DSBs when the uracils are less dense, by EXO1-mediated degradation of DNA regions between two BER-created nicks, resulting in a DSB [152, 247]. Although widely-accepted, this mechanism has never been characterized biochemically. With the present study, we attempt to fill this gap and investigate the mechanism of DSB induction in detail.

EPIGENETICS

Epigenetics entails a variety of processes that modify DNA 'in addition' (Greek: *epi-:* over, above, in addition to, around) to the base sequence and chromatin structure, leading to transient or stable changes in gene expression.

In addition to the direct modification of DNA with covalently-attached methyl groups on cytosines (or adenines), also the building blocks of nucleosomes, the histones, can become altered by a large variety of post-translational modifications (PTMs). DNA- and histone modifications affect the structure of chromatin not only directly, but also indirectly, through bringing about alterations of

interactions with other factors. Which of these processes plays the primary role in the modulation of gene expression has been the subject of intense debate during the past several decades. Mounting evidence points towards a complex interplay that depends on the stimulus and the type and developmental stage of the cell.

In eukaryotes, genomes are packaged into two main types of chromatin: the relaxed or "open" euchromatin and the densely-packed heterochromatin. The differences are established by a combination of DNA methylation and the type and modifications of the associated proteins. Actively-transcribed genes are generally located in euchromatic regions, whereas other genomic regions, containing mostly inactive genes, are packed into heterochromatin. The latter can be divided into two different types: facultative and constitutive. An example of facultative heterochromatin is the inactive X-chromosome, marked extensively with histone H3 trimethylated on lysine 27 (H3K27me3) and the Polycomb Repressor Complex (PRC), responsible for transcriptional silencing during development [248]. Constitutive heterochromatin contains permanently-silenced genes. Examples are chromatin of the centromeres or telomeres that are characterized by high abundance of H3K9 trimethylation and binding of heterochromatin protein 1 (HP1) that recruit other chromatin remodeling proteins and DNA methyltransferases, leading to DNA compaction [249].

Increasing attention in the field of epigenetics is given to different kinds of RNAs that influence transcription, among others long-noncoding RNAs and micro RNAs (reviewed in [250-253]). This topic will not be addressed here.

Histone modifications

Eukaryotic DNA is organized and packed into nucleosomes, which consist of five main histones in eukaryotic cells: H1, H2A, H2B, H3 and H4 and several minor specialized histone variants. Two H2A-H2B dimers and one H3-H4 tetramer form the nucleosome core that is stabilized by the linker histone H1 [254]. A diverse array of modifications of the histones controls chromatin organization.

The two main and largely opposing histone modifications are acetylation and methylation of lysines and arginines on the histone N-terminal tails. Acetylation of lysines has a positive effect on RNA synthesis [255] by largely neutralizing the positive charge of the histone, thereby weakening the interaction with the negatively-charged DNA. The resulting less-tightly packed chromatin is more accessible to the transcription machinery. Histone acetylation is regulated by histone acetyltransferases (HATs) and deacetylases (HDACs). Acetylated lysines are bound by bromodomains, often found in chromatin-remodeling complexes that function to open the chromatin structure [256].

Methylation of histones is more complex, as there are different combinations of modifications. Lysines for example can be mono-, di- or tri-methylated and depending on the histone type and the site modified, methylation can lead to activation or repression of transcription. Methylation patterns on histones are established by histone methyltransferases that use S-adenosyl methionine (SAM) as the methyl group donor and removed by histone demethylases. Specialized domains of chromatin-associated factors can bind methylated lysines, among others zinc finger plant homeodomains (PHDs), chromodomains, PWWP (proline-tryptophan-tryptophan-proline) motifs, Tudor or MBT

(malignant brain tumor) domains [257]. Some methylation marks on histone 3 that are especially important for gene regulation will be discussed in more detail.

A histone mark strongly linked to active genes is H3K4 methylation. Enhancers responsible for active transcription contain high levels of H3K4me1 [258], whereas transcriptional start sites (TSS) of active genes are marked by enrichment of H3K4me3 [259, 260].

H3K4me3 is bound by tandem chromodomains within chromodomain-helicase-DNA-binding protein 1 (CHD1), an ATP-dependent remodeling enzyme capable of repositioning nucleosomes [261]. In addition, H3K4me3 prevents binding of the nucleosome remodeling and deacetylase (NuRD) repressor complex to the H3 N-terminal tail [262]. Both these interactions lead to chromatin opening. H3K36me3 is also associated with active chromatin, but it is highly enriched throughout the entire transcribed region [263].

H3K27 methylation, the mark of facultative heterochromatin as mentioned above, is established by the PRC, specifically by the lysine methyltransferase EZH2, leading to compaction of chromatin during development, for example on the inactive X chromosome [264].

H3K9me3 is bound by HP1 and also correlates with repressive chromatin environment. HP1 binds to H3K9me3 *via* its N-terminal chromodomain and this interaction is important for the overall structure of heterochromatin [265, 266]. HP1 has also been shown to recruit the DNA methyltransferase 3B (DNMT3B), thereby linking histone and DNA methylation [267]. DNA methylation in contrast can also inhibit protein binding to specific histone modifications. An example is lysine demethylase 2A (KDM2A) that is only able to bind nucleosomes bearing H3K9me3 when the DNA is unmethylated [268].

The two methylation systems interact to form stable heterochromatic subdomains and thereby protect genome integrity.

To make the story even more complex, histones can also be ADP ribosylated, phosphorylated, ubiquitylated, or SUMOylated, enabling the recruitment of a large variety of modulating proteins (reviewed in [269, 270]).

DNA methylation

The modification of adenines and cytosines with a covalently-attached methyl group can alter protein/DNA interactions. In bacteria, this forms the basis of restriction/modification systems; a primitive immune system in which invading DNA molecules (e.g. phage genomes) are degraded by nucleases that cleave unmethylated DNA.

This phenomenon was first reported by Werner Arber in the early 1960s. He hypothesized that enzymes in bacteria cut foreign DNA into pieces at defined sequences and that this is impeded by methylases of the host [271-277]. Confirmation of this process came from Hamilton Smith, who discovered the endonuclease *Hind*II in 1969. He showed that cleavage depended on a specific sequence of 6 base pairs [278, 279]. Daniel Nathans then used Smith's and others' restriction

enzymes to create a map of simian virus 40 (SV40), by cutting it into defined pieces, applying restriction enzymes to genetics for the first time [280-285]. The discovery of this restriction/modification system opened a whole new field of molecular biology and helped to solve various problems in genetics, not only in bacteria, but also in eukaryotes.

Apart from host protection, methylation in bacteria was shown to have another important purpose, namely to act as a strand-discrimination signal in MMR [121, 286-289]. *E.coli* strains either lacking or overexpressing the *Dam* methylase responsible for post-replicative methylation of GATC motifs exhibit a hypermutator phenotype [288, 290-293], similar to MMR mutants [294, 295].

DNA methylase activity is also present in mammals [296, 297], where it serves different purposes. While the main modified bases in bacteria are both, 6-methyladenine and 5-methylcytosine, only the latter plays a role in higher eukaryotes. In plants and vertebrates, DNA methylation of cytosines plays a key role in development through controlling gene expression. Physiological DNA methylation occurs in CpG, CpHpH or CpHpG (H describes a non-guanine residue) sequence context in plants [298, 299], but mainly in CpG dinucleotides in vertebrates [300]. This work will focus on DNA methylation in vertebrates, where the majority of CpGs (70-90%) are methylated [301].

Nearest-neighbor sequencing studies revealed many years ago that the frequency of CpG dinucleotides in DNA is much lower than expected from random distribution [302]. This phenomenon was suggested to be linked to DNA methylation, as there was a correlation between deficiency in CpGs and high methylation levels in different species. An explanation came from studies in *E. coli*, where sites of cytosine methylation were shown to be hotspots of C to T transition mutations, believed to be associated with spontaneous deaminations of 5mC to T that are not repaired [303]. To confirm this phenomenon, TpG and CpA dinucleotide frequencies were assessed and found, as expected, to be above average in frequency [304]. Several groups set out to calculate the mutation rate of methylated CpGs. Depending on the gene investigated and the exact formula applied, values from 10 to 42 times higher mutation rates at 5mC than at other sites were found [305-307]. Several intriguing questions arose from these findings, concerning the role and the consequences of 5mC and its mutability. Are there repair mechanisms counteracting the high mutation frequency? Might this phenomenon be just a way to increase mutation rate beneficial for evolution? Or is there an important function of 5mC, outweighing the disadvantage of high mutability?

In an attempt to identify repair pathways acting on 5mC deamination that results in G:T mismatches, the MMR protein MutSα [122, 123] and the BER enzyme TDG [73, 75] were found, as discussed above.

A role of 5mC mutability in evolution seemed unlikely, because it brings about only a limited number of amino acid changes and would gradually lose significance through the progressive loss of mutable sites. The notion that cytosine methylation played an important biological role outside of evolution, counterbalancing its mutability, gained support with the discovery of CpG islands.

Apart from a significant underrepresentation of the CpG dinucleotide, an asymmetrical distribution of CpGs was found [308]. In the 5' region of many genes there was an accumulation of CpGs. Due to this pattern, a role for 5mC in gene expression was postulated [309]. A study in Chinese hamster

confirmed this hypothesis. Two housekeeping genes, possessing dense CpG regions at the 5' region, were analyzed using methylation-sensitive restriction enzymes and found to be unmethylated at the CpGs. *In vitro* methylation caused the silencing of the genes, directly linking methylation to gene expression [310]. Similar findings were obtained from a study on human gamma-globin gene expression [311]. Further evidence for the inhibitory effect of methylation at CpG-islands came from studies of the hypoxanthine phosphoribosyltransferase (HPRT) gene on the X chromosome. It was found to be inactive when methylated and could be induced with 5-azacytidine, a cytosine analogue whose incorporation into DNA leads to demethylation [312-314].

In the following years, more examples of primarily housekeeping genes with unmethylated CpG islands in the 5' region were found [315, 316]. The generally unmethylated state of these regions explains the accumulation of CpGs, as unmethylated cytosines are more resistant to deamination and if still deaminated, extremely efficiently repaired by BER, which prevents their loss from DNA.

Studies on the X chromosome [317] and on imprinted genes [318, 319] showed however that CpG islands are not inherently resistant to methylation. Moreover, many cancers exhibit CpG hypermethylation of tumor suppressor genes [320, 321]. One prominent example is aberrant methylation of the *MLH1* promoter, leading to inactivation of the gene and tumorigenesis in endometrium and colon [322-325]. It is therefore of interest to investigate the mechanism by which CpG islands are selectively kept free of methylation. Several hypotheses are under consideration. There is evidence that the *de novo* methylation capacity of adult tissues is very weak [326]. This would guarantee the maintenance of unmethylated sites after methylation patterns have been established during development. Another idea is that methylation is a secondary consequence of gene inactivity. As long as a gene is active, its CpG island might be bound by transcription factors or other accessory proteins, preventing access for methylases. There is accumulating evidence for proteins binding unmethylated CpG islands, therefore protecting them from methylation or even constantly demethylating them. Accordingly, the TET1 protein has been shown to localise to CpG islands via its CXXC domain [327, 328] and depletion of TET1 resulted in increased CpG methylation [329]. In addition, experiments in TDG knockout mice suggested that the glycosylase may play a role in protecting CpG islands from *de novo* methylation [39, 98]. These pathways will be discussed in more detail later.

The study of CpG islands received a great deal of attention, but also single CpGs have been linked to gene silencing and shown to be modified dynamically. One example is the thymidine kinase promoter, where one CpG was enough to significantly downregulate transcription when methylated [330]. In this thesis, the focus lies on the vitellogenin system, discussed later, where methylation of few CpGs impaired transcription [331, 332].

DNA methylation and chromatin structure

The effect of DNA methylation on gene expression was found to be mediated by altered DNA-protein interactions leading to changes in chromatin structure as mentioned already in 'histone modifications'. Studies using DNaseI and micrococcal nuclease (MNase) showed that unmethylated DNA stretches assume a DNase I-sensitive conformation that was prevented by CpG methylation,

rendering the sequence DNase I-insensitive. These findings led to the conclusion that DNA methylation modifies specific, as well as nonspecific interactions between DNA and proteins [333]. Support for these findings came from studies of an upstream region of the tyrosine aminotransferase (TAT) gene, where methylation prevented specific factor binding in cells that do not express TAT [334]. Another example is the methylation of the cAMP-responsive element that was shown to suffice to abolish transcriptional activation by inhibiting transcription factor binding [335]. In the following years, many transcription factors were found to be prevented from binding DNA when their recognition site was methylated [336-341]. Specific transcription factor binding often initiates the changes in chromatin leading to the establishment of a transcription-friendly environment. Inhibiting their binding therefore prevents recruitment of chromatin remodelers, impeding transcription.

The discovery of a group of proteins binding specifically to methylated CpGs added another layer of complexity to the control of chromatin dynamics by DNA methylation. The first methyl-CpG binding domain (MBD) protein discovered was MeCP1 [342]. It was shown to bind DNA stretches with an accumulation of symmetrically-methylated CpGs, leading to inhibition of transcription [343]. MeCP2, detected shortly thereafter, displayed similar properties, but was able to bind a single 5mCpG pair *in vitro* [344]. *In vivo* it was shown to localize to centromeric heterochromatin in a methylation-dependent manner [345]. It has an important role in embryonic development, but was dispensable in embryonic stem cells (ESCs) [346]. MeCP2 binding leads to repression of transcription *in vitro* as well as *in vivo*, displaying a direct correlation with the density of 5mCpGs [347]. Three additional members of the MBD family were found in a database search for proteins with an MBD similar to MeCP1 and 2, as mentioned earlier [99]. MBD proteins recruit chromatin remodelers, histone deacetylases and DNA methyltransferases, leading to compaction of chromatin [348, 349]. Deletions of MBD1, MBD2 and MBD4 show only mild developmental phenotypes, pointing towards redundant functions. *Mbd3*$^{-/-}$ mice are embryonic lethal, most likely due to the important role of MDB3 as part of the NuRD repressor complex [350, 351]. Mutations of MeCP2, however, are mostly lethal in males and cause Rett Syndrome, a neurological disorder, in females [352-354]. Rett syndrome is not a neurodegenerative disorder, accordingly, symptoms have been successfully reversed in mutant mice by re-expression of the protein [355].

DNA methylation acts through recruitment of a combination of unspecific and specific interacting proteins. MBD proteins play a major role in silencing regions with a high density of methylated CpGs, such as methylated CpG islands. Cell type-specific genes, however, often have only few CpGs in their regulatory regions. In that case, specific transcription factor binding seems to be inhibited by DNA methylation. These events often occur in enhancers instead of promoters, complicating the analysis.

Chromatin and transcription

The compaction of DNA into chromatin, guided by histone and DNA modifications as mentioned above, is necessary to fit the large DNA molecule into the limited space of the nucleus. It brings about a flexible and versatile way of controlling gene expression, but also a number of challenges, due to limited accessibility of the DNA.

Gene expression is controlled primarily by regulatory DNA elements and chromatin structure. The minimal elements defining a gene on DNA are a promoter region, containing sequence motifs that bind transcription factors (TFs), and a transcription start site (TSS), defined by these sequence motifs. Most genes have additional regulatory elements, called enhancers, often distal from the gene itself. Promoters can be broadly divided into CpG-poor and CpG-rich, the latter containing CpG islands, as discussed in 'DNA methylation'. Most, if not all, housekeeping genes are associated with CpG islands and therefore 'open' chromatin, whereas only around half of tissue-specific genes display this feature. A ubiquitous TF that binds the consensus sequence GGGCGG (or CCGCCC), a sequence prominent in CpG islands, is Specificity Protein 1 (SP1). Most, but not all promoters additionally possess a TATA-box ~30 nt upstream of the TSS that is bound by a TATA-box binding (TBP) protein.

Transcription factors are responsible for communicating the gene-specific DNA sequence information to the transcription machinery. They consist of two main domains: a DNA binding domain that "reads" the binding site sequence, and a transactivation domain. The DNA binding domain can assume different forms. One group of TFs possesses helix-turn-helix motifs consisting of two α-helices connected by a short amino-acid stretch. The C-terminal helix is placed in the major groove of the DNA, where it can make base-specific contacts. The members of this group bind DNA as dimers. Other DNA binding proteins contain so-called zinc fingers, domains that coordinate one or more zinc ion to stabilize their structure [356]. Leucine zippers constitute a special case of DNA binding domains, as they mediate - in addition to DNA binding - also protein-protein interactions *via* extended contacts along their α-helices [357, 358]. A related DNA-binding motif is the helix-loop-helix (HLH) structure. It also mediates dimerization in addition to DNA binding. A HLH motif is composed of a short α-helix connected to a longer one *via* a flexible loop [359].

The different TF motifs “read” specific nucleotide sequences, primarily through hydrophobic interactions in the major grove of DNA, but also with the help of base-specific interactions defined by combinations of hydrogen bond donor and acceptor sites. Some TF recognize their cognate DNA sequences with the help of other characteristics, such as base-stacking or helical twist. Depending on the flexibility of the DNA stretch, DNA binding proteins can cause distortions of the helix, enabling them better access. Some TFs show moreover sensitivity to cytosine methylation in their recognition sites, as mentioned above, while other DNA binding factors, containing MDB domains, bind specifically to methylated DNA as transcriptional repressors [360-363].

The other domain defining a transcription factor can also assume different forms. Generally speaking, the activation domain mediates the effector function of the TF. This happens *via* interactions with other proteins, for example RNA polymerase II. This represents the most direct way of activating transcription. Many TF also act *via* chromatin remodeling interactors, thereby preparing the chromatin environment for gene expression. There are three main motifs characteristic of TF activation domains: acidic-, glutamine-rich- and proline-rich motifs. Acidic activation domains were shown to assemble in an amphipathic helical structure, where one side contains the acidic, hydrophilic residues, whereas the other side is hydrophobic [364]. Glutamine-rich domains have been most extensively studied in SP1, where they mediate the interaction with TBP, a subunit of TFIID [365-367]. Regardless of whether the activation domain of a TF interacts with other TFs or directly with RNA polymerase, the ultimate goal is recruitment of the latter enzyme to the TSS.

Eukaryotes possess three different nuclear RNA polymerases that are responsible for the transcription of distinct classes of genes [368]. The most extensively-studied is RNA polymerase II, which transcribes all protein-coding genes into messenger RNAs (mRNAs) and a variety of non-coding RNAs. RNA polymerase I is responsible for the synthesis of the long ribosomal RNAs (rRNAs) and polymerase III for transcription of the small rRNA and transfer RNAs [369]. The three enzymes consist of several subunits, some of which are shared, resulting in common characteristics. In contrast to the prokaryotic enzyme, eukaryotic RNA polymerases rely on TFs for the initiation of transcription. The factors that are absolutely needed to enable basic transcriptional activity and are involved in every transcription event, are called general TFs. They help recruit the transcription machinery, consisting of an RNA polymerase and chromatin remodelers. There are five general TFs needed for transcription of protein-coding genes by the eukaryotic RNA polymerase II. The first step is the binding of TFIID (containing TBP) to the TATA-box. TBP was shown to have a relatively broad sequence specificity and binds also to promoters lacking a TATA-box. TBP then binds TFIIB and together they recruit TFIIF and RNA polymerase II. Then, TFIIE and H, important for initiation of transcription, can bind. TFIIH is a multi-subunit protein with several enzymatic activities. It comprises helicases to open DNA for transcription and a kinase that phosphorylates the C-terminal domain (CTD) of the largest RNA polymerase subunit [370]. The CTD has many essential functions in transcription, not only in initiation, but also in elongation, processing of mRNAs and in termination. It consists of up to 52 copies of the heptad repeat Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Phosphorylation of CTD by the kinase subunit of TFIIH helps to release the polymerase from the transcription initiation complex, enabling it to start elongating the RNA. A complex interplay of different kinases and phosphatases modifies the CTD while it travels along the gene, enabling efficient transcription, as well as orchestrating 5'-capping, splicing and 3'-polyadenylation [371].

This complex general mechanism of gene expression is further modulated and fine-tuned by many additional factors, especially in the case of tissue-specific genes, but also in ubiquitously expressed housekeeping genes.

To account for the different tissue- and developmental stage-specific expression patterns in the cells of higher eukaryotes, a wide array of additional TFs is needed. Gene-specific transcriptional regulation is generally governed by the binding of transcriptional activators or repressors to upstream regulatory sequences, the so-called enhancers. They recruit chromatin-modifying enzymes, preparing a transcription-permissive environment for the general TFs and the RNA polymerase [372]. Often, these tissue-specific transcription factors react to intra- or extracellular stimuli, such as cellular metabolites, interleukins or hormones. They are thereby able to merge a variety of signals, translating them into cell type-specific gene expression. Hormone-dependent transcription factors belonging to the superfamily of nuclear receptors, will be discussed in more detail below.

DNA methyltransferases

DNA methyltransferases (DNMTs) are a highly conserved family of proteins. In jawed vertebrates, there are three major DNMTs displaying robust methyltransferase activity: DNMT1, DNMT3A and DNMT3B. DNMT2 has only weak enzymatic activity and is not well described [373]. Another member

of the family, DNMT3L [374], has no enzymatic activity, but was shown to increase activity of the other DNMT3s [375, 376].

The C-terminal catalytic domain of the DNMTs consists of 10 conserved amino acids forming a prolyl-cysteinyl active site that catalyzes the attachment of a methyl group to the 5-position of cytosine through a complex process first described in bacterial homologs of the DNMTs [377-379]. A conserved cysteine residue initiates the reaction by attacking the carbon at the 6-position, forming an covalent intermediate between the cytosine and the enzyme (Fig 12). This produces a negative charge at position 5 and activates the usually inert carbon that can now interact with the methyl donor S-adenosyl methionine (SAM). The product, S-adenosyl-homocysteine (SAH), is released and the proton at position 5 removed by a basic amino acid on the enzyme. A β-elimination step completes the reaction, releasing an active enzyme and 5mC that returns into the double helix [380-382].

The demethylating agent 5-aza-deoxycytidine (5-AzaC) is a cytosine analog that bears a nitrogen at position 5, instead of a carbon. This leads to the formation of a covalent bond with the enzyme that cannot be β -eliminated [377, 378], thereby trapping the DNMT on the DNA.

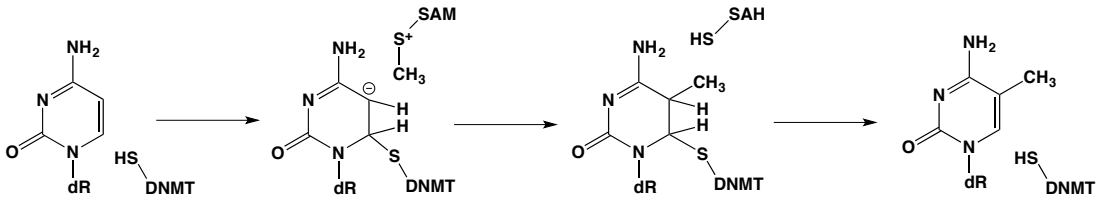


Figure 12 Mechanism of cytosine methylation in DNA, catalyzed by DNA methyltransferases.

The first mammalian relative of bacterial restriction methyltransferases, DNMT1, was cloned in 1988 [383]. It was shown to have 5- to 30-fold preference for hemimethylated substrates, strongly implicating it in maintenance of methylation during DNA replication [384]. It shares the C-terminal catalytic domain with DNMT3A and B, but has an additional, large N-terminal domain, implicated in regulation and targeting. The interaction with Ubiquitin Like With PHD And Ring Finger Domains 1 (UHRF1, also nuclear protein 95) mediates the selectivity for hemimethylated site recognition, *via* the SRA (SET and RING finger associated) domain of UHRF1 [385-387]. In addition, DNMT1 contains a zinc finger CXXC domain that is characteristic of proteins binding unmethylated CpGs. Other proteins containing this domain include the lysine demethylases KDM2A/B and the ten-eleven-translocation (TET) proteins 1 and 3. DNMT1 also possesses a PCNA interacting peptide (PIP) module responsible for targeting the protein to replication foci during S phase. This interaction has been shown to increase the enzyme efficiency two-fold, but the recruitment to the replication fork is not absolutely necessary to maintain global methylation levels [388, 389].

The N-terminal domain of DNMT1 was shown to inhibit *de novo* methylation activity of the C-term [390-392], further increasing the specificity towards maintenance methylation.

Evidence for the existence of further DNA methyltransferases came from knockout studies of DNMT1. *Dnmt1*$^{-/-}$ embryos die during development, but ESCs grow normally and retain basal levels of

methylation [393], pointing towards a backup system for methylation of cytosine. *Dnmt1*$^{-/-}$ ESCs undergo apoptosis when induced to differentiate however, leading to the conclusion that the backup system is specific for embryonic development [394]. A database search for 5-cytosine methyltransferase motifs performed following these finding revealed the existence of DNMT3A and B [326]. Apart from the C-terminal catalytic domain, they show little sequence similarity to DNMT1. They are highly expressed in undifferentiated ESCs. DNMT3A levels in adult tissues are very low and DNMT3B is barely detectable. They show robust methyltransferase activity on hemi- as well as unmethylated DNA, assigning them a role as *de novo* DNA methyltransferases [395].

While DNMT1 is uniformly expressed in all cell types and developmental stages, DNMT3A and B show distinct cellular localizations during embryogenesis hinting towards separate roles. Like DNMT1, knockouts of DNMT3A and B lead to early lethality. *Dnmt1*$^{-/-}$ and *Dnmt3B*$^{-/-}$ do not survive the embryonic stage, whereas *Dnmt3A*$^{-/-}$ mice appear normal at birth, but die at around 4 weeks of age [396].

DNMT3B is required for *de novo* methylation of pericentromeric repetitive DNA sequences and CpG islands on the inactive X chromosome. Mutations in both alleles of DNMT3B lead to the development of the ICF (immunodeficiency, centromeric region instability, and facial anomalies) syndrome in humans [397-400]. Consistent with the fact that DNMT3B loss is embryonic lethal, it was found that at least one mutation in ICF patients has residual catalytic activity [401]. One hallmark of this genetic disorder are low methylation levels in pericentromeric satellite regions [402] and the disease translates into genomic instability and fatal immunodeficiency.

Deregulation of all major DNMTs is associated with cancer development [403, 404], assigning them a role as potential anti-cancer targets. Methyltransferase inhibitors that competitively bind to the catalytic site of DNMT have been used in clinical trials to treat different kinds of hematopoietic malignancies. 5-AzaC (Decitabine, trade name: Dacogen) is approved for the treatment of myelodysplasic syndromes, including different kinds of anemias and chronic myelomonocytic leukemia [405-408].

DNA demethylation

DNA demethylation can be divided into two major mechanisms: passive demethylation, involving replication-dependent dilution of the methyl marks, and active demethylation that is replication-independent and needs the action of specialized proteins. The most studied demethylation events occur in the paternal pronucleus after fertilization [409, 410] and during primordial germ cell (PGC) development [411, 412]. Both were described to involve active demethylation events, as they occur rapidly, before DNA replication takes place.

Most studies to date described genome-wide changes in methylation during development, involving CpG-rich regions, such as CpG islands. The sensitivity of the applied methods does not allow conclusions on specific single CpGs. As mentioned earlier, however, demethylation of single CpGs can also have a profound effect on expression of certain genes in somatic cells. The proteins involved in the drastic demethylation events occurring during development and the subtle changes in

methylation in somatic cells are not clearly defined. There are several hypotheses under discussion that will be presented briefly.

The direct removal of the methyl group would be the simplest mechanism of demethylation and has been shown to occur during the repair of 6-*O*-methylguanine where MGMT removes the O-linked methylation on damaged DNA, thereby irreversibly inactivating itself. But, in contrast to the latter alkylated base, the methyl group on 5mC is attached via a carbon-carbon bond. Breakage of this bond is thermodynamically very unfavorable, rendering this mechanism unlikely.

If the damage cannot be reversed directly, the next idea would be to remove the entire nucleotide. Enzymes capable of carrying out this reaction have been found in plants. The Demeter/repressor of silencing 1 (ROS1) family of glycosylase/lyases can initiate the active demethylation process by excising 5mC, leaving behind an incised abasic site that can be repaired by BER [413]. Although described in several reports [93, 103, 414, 415], no enzyme with robust glycosylase activity towards 5mC has actually been purified from animal cells.

Another highly-debated hypothesis involves enzymatic deamination of 5mC to T. It is a very tempting model, because - as already described - there are several possibilities to repair the resulting G:T mismatches, leading to demethylation. The main player described for this pathway is AID. Although mainly expressed in activated B-cells, low levels of AID have been detected among others in primordial germ cells [202], which undergo a wave of rapid, active demethylation. The cytidine deaminase displays activity on 5mC, albeit much weaker than on unmethylated cytosine [235, 236]. This raises the question of how the cell would protect the cytosines in an environment with high levels of AID activity. Despite this caveat, genome-wide studies found a correlation between AID activity and demethylation in PGCs [237, 238, 416] and in embryogenesis [98, 104, 417].

With the discovery of the Ten-Eleven-Translocation (TET) enzymes, the field of DNA demethylation got another interesting twist. It proposes a way of demethylation involving oxidation of 5mC, instead of deamination, and the excision of the oxidation products by a glycosylase. It is the most supported hypothesis at the moment and will be discussed in more detail.

Ten eleven translocation proteins in demethylation

Ten-eleven translocation (TET) proteins constitute a new family of Fe^{2+} and 2-oxoglutarate (2OG) dependent dioxygenases. Their existence was first mentioned in 2002, when a translocation of the H3K4 methyltransferase mixed lineage leukemia (*MLL*) gene on chromosome 11 to a gene on chromosome 10 encoding a protein with a CXXC domain was observed in acute myeloid leukemia (AML) [418, 419].

It was only several years later that the enzyme was purified and characterized. It was shown to be one of three enzymes in mammals that catalyze the oxidation of 5mC to 5hmC [28] and further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [27] *in vitro* and in cells. 5fC and 5caC are substrates for TDG, providing an attractive model for sequential demethylation [97, 420]. All five cytosine species are known to be present in numerous mammalian tissues. 5hmC is present at 1-5%

of 5mC in most-cell types [421, 422], up to 10% in ESCs [28] and as high as 40% in Purkinje neurons [423]. 5fC and 5caC levels are much lower (0.03 and 0.01% respectively) compared to 5mC in ESCs [27, 97].

The family of TET dioxygenases is widely distributed among eukaryotes, from the amoeboflagellate Naegleria gruberi to mammals. In jawed vertebrates, the *TET* gene underwent triplication, giving rise to *TET1*, *2* and *3*. Chromosomal inversion of *TET2* detached the CXXC domain, which became an independent gene that is translated into *IDAX* (inhibition of the Dvl and axin complex, also known as *CXXC4*).

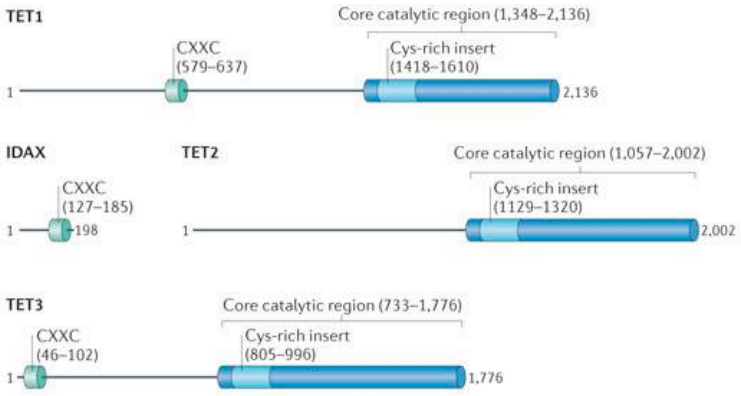


Figure 13 Schematic representation of the three TET proteins present in jawed vertebrates. They share a highly conserved catalytic region at their C-terminus. The TET2 locus underwent inversion, resulting in the detachment of the DNA binding domain CXXC. From [424].

The core catalytic domain of the TET enzymes forms a double-stranded β-helix (DSBH) fold, characteristic of Fe^{2+} -and 2OG-dependent dioxygenases. It contains a triad of two histidines and one aspartate or glutamate that cooperate to bind Fe^{2+}. The iron center then binds oxygen and 2OG, together with a conserved arginine and a serine residue. TET enzymes are dioxygenases, because they incorporate both atoms of molecular oxygen, coordinated by the iron, into their products; one into the co-factor 2OG, resulting in the release of CO_2 and succinate, and one into 5mC, 5hmC or 5fC [425-428]. Ascorbate was shown to enhance TET catalytic activity [429], likely by acting as a reducing agent to return the metal center iron to its active state [430, 431].

Studies on a *Naegleria* TET-like dioxygenase (*NgTET1*) reveal that 5mC is strongly preferred as substrate over 5hmC and 5fC. It was further shown that more 5caC was formed when the initial substrate was 5mC, then when the reaction was started with DNA containing 5hmC or 5fC. This led the authors to the conclusion that the three oxidation steps might be carried out processively by TET, without releasing the DNA [432, 433]. After every oxidation step, however, succinate has to be released to make space for a new 2OG. Structural studies on human TET2 revealed that the hydrogen abstraction leading to the oxidation of the cytosine species is likely to be inefficient from the hydroxyl and carbonyl groups of 5hmC and 5fC. This results in low catalytic efficiency [434]. The

crystal structures of *NgTET1* and TET2 revealed a base flipping mechanism to carry out the oxidation of their substrates, indicating that only one cytosine in a symmetrical CpG is addressed at a time.

TET1 and 3 contain a CXXC motif in the N-terminal domain, known from other DNA binding proteins and implicated in binding of unmethylated CpG-rich regions [435, 436]. Metazoan TETs additionally possess a cysteine (cys)-rich domain in the catalytic region at the C-terminus, adjacent to the DSBH domain. The cys-rich region chelates two or more zinc ions and is postulated to bind DNA and help in target recognition. Crystal structure of a TET2 complex with methylated DNA showed that the cys-rich domain wraps around the DSBH and together they coordinate 3 zinc ions thus stabilizing important structures for catalysis and DNA binding [437].

The different TETs in jawed vertebrates carry out the same catalytic reaction, but are expressed in different tissues and developmental stages.

TET1 levels are barely detectable in early zygotic stages, but increase at later pre-implantation stages and are high in ESCs, together with TET2 [438, 439]. There is limited expression of TET1 in several adult tissues [418]. TET1 localizes predominantly to CpG islands, keeping them methylation-free, presumably by inhibiting *de novo* methylation by DNMTs or by active demethylation. Depletion of TET1 leads to increased levels of 5mC at many TET1-enriched regions. These experiments also revealed a dual role of TET1 in gene regulation, as similar number of genes were upregulated and downregulated in the absence of TET1. This points towards a complex interplay of different factors and complicates the analysis on a genome-wide scale [329, 440, 441].

Tet1 knockout mESCs show a decrease in 5hmC levels of approximately 35%, but maintain normal morphology and support embryonic development. *Tet1*-null mice are viable, but the number of progenies is reduced indicating that either a fraction of embryos die or germ cell development is impaired [442]. *Tet1*- and -*2* double knockout mice are also viable and develop all three germ layers, indicating that the two TETs have modulatory, but not essential functions in development [443].

TET1 and 2 have been implicated in the rapid demethylation event occurring in PGCs. The demethylation is accompanied by conversion of 5mC to 5hmC [444, 445]. Kinetics of removal of 5mC and 5hmC indicate a passive mechanism involving DNA replication, after rapid conversion of most of the methylation to hydroxymethylation. TET1 and TET2 therefore seem to enhance passive demethylation, but they are not the only factors responsible, in agreement with the largely normal phenotypes of the surviving progenies from $Tet1^{-/-}$ $Tet2^{-/-}$ mice [443].

TET2 levels are, similarly to TET1, high in later stage embryos and in ESCs. TET2 is additionally ubiquitously expressed in adult tissues [418], including high levels in hematopoietic stem cells (HSCs). Accordingly, loss of TET2 was observed to promote HSC self-renewal, proliferation, and differentiation, leading to a number of hematopoietic malignancies [446-449]. TET2 displays haploinsufficieny, as already $TET2^{-/+}$ HSCs show a growth advantage for populating hematopoietic lineages [450]. Although there are also low levels of TET1- and- 3 expressed in the hematopoietic system, loss of TET2 alone causes significant loss of 5hmC and results in tumorigenesis, assigning it a role as tumor suppressor in HSCs [446]. As TET2 lacks a CXXC DNA binding domain, its targeting to DNA is likely to be mediated differently from TET1 and 3, probably by interacting proteins.

TET3 is enriched in oocytes and early zygotes [451]. It has been postulated to be responsible for the rapid demethylation events taking place after fertilization in the paternal pronucleus. Appearance of 5hmC coincides with the disappearance of 5mC from the paternal pronucleus. [439], where TET3 appears to be specifically enriched [452]. Further evidence for this pathway was provided by knockdown studies showing impeded conversion of 5mC to 5hmC in the absence of TET3 [439]. In a recent study, however, it was found that the initial loss of paternal 5mC does not require TET3-dependent 5hmC formation [453], pointing towards a more complex mechanism likely involving other, yet unknown, factors. In addition, the observation was made that 5hmC and, to a lesser extent, 5fC and 5caC, persist in embryos and only gradually decline until the eight-cell stage [438, 454, 455], consistent with passive dilution during replication. The intriguing question arises how the TET activity is halted mostly at 5hmC. The oxidation of 5mC by TET3 nevertheless has an important role, as mice produced from TET3 knockout oocytes are not viable, regardless of the genotype of the male [452].

Apart from the role of the TET enzymes in active DNA demethylation, its oxidized products themselves might also play important roles in DNA. One well-established example concerns maintenance methylation. DNMT1 strongly prefers hemi-methylated- over hemi-hydroxymethylated sites [456]. This results in passive, replication-coupled demethylation, following a first active enzymatic step. Also the MBD proteins were shown to have reduced binding affinity towards 5hmC, as compared to 5mC [457]. The binding properties of specific transcription factors, such as hormone receptors, towards the oxidized cytosine species remain to be analyzed and will provide further information about the regulatory capabilities of the TETs and their products.

Thymine DNA glycosylase in demethylation

TDG was first implicated in chromatin dynamics and transcriptional control *via* its interaction with hormone receptors [458-460] and the histone acetyltransferase CBP/p300 [461]. Knockouts of both CBP [462] and p300 [463] display similar defects as TDG knockout mice, indicating that TDG knockout lethality is connected to chromatin dynamics, rather than its canonical repair activity.

The first evidence that TDG might play a role in DNA demethylation was obtained when it was purified from a chicken 'demethylation complex' and shown to have 5mC glycosylase activity [93]. This activity was however extremely low compared to the activity against G/T mismatches and therefore most likely not physiologically relevant, unless unknown auxiliary factors would increase the activity substantially.

The connection between TDG and demethylation, however, received further support in the following years. Overexpression of TDG in human kidney cells lead to demethylation of a hormone-responsive reporter gene that was necessary for its expression [460]. TDG was also linked to myoblast differentiation that was accompanied by genome-wide DNA demethylation [464]. In a functional genomics screen aimed to identify proteins involved in demethylation of a reporter gene, the SUMO-dependent ubiquitin E3-ligase RING finger protein 4 (RNF4) was found [465]. RNF4 deficiency is embryonic lethal and *Rnf4*$^{-/-}$ mouse embryonic fibroblasts (MEFs) present higher levels of 5mC. RNF4 was shown to interact with TDG and APE1 and the interaction enhanced demethylation of the

reporter. RNF4 increased the enzymatic activities of TDG and APE1 as shown on a G/T mismatch. The authors postulated an involvement of a deaminase, but detected no involvement of AID or APOBEC1. Another study saw no RNF4-mediated demethylation, but found that RNF4 is responsible for MeCP2 turnover, thereby activating transcription of methylated genes [466].

Tdg knockout mice, shown to be embryonic lethal, shed more light on the role of TDG in methylation dynamics. TDG was shown to protect unmethylated regions from aberrant *de novo* methylation, as *Tdg*$^{-/-}$ MEFs display hypermethylation at regions normally unmethylated and bound by TDG [39]. Another group that also established a *TDG* knockout model showed that two site-specific demethylation events described earlier [467, 468] are dependent on TDG. They proposed a pathway involving AID and GADD45a (Growth Arrest and DNA Damage-inducible 45) [98], based on the fact that AID interacts with TDG and GADD45a. The involvement of GADD45a in DNA demethylation is controversial. Some groups support its role [104, 469], whereas others do not [470].

With the discovery of the TET enzymes and the ability of TDG to excise their oxidation products, the connection between TDG and DNA demethylation was further strengthened [27, 28, 97, 420]. 5fC and 5caC accumulate in proximal and distal regulatory elements in mESCs upon depletion of TDG [471] and a connection between 5fC, TDG and the p300 histone acetyltransferase was found at enhancers [472]. Recombinant TET1 and TDG were shown to interact physically and to process 5mC and 5hmC *in vitro*, resulting in alkaline labile AP-sites [95].

Despite extensive research on the pathway of active DNA demethylation *via* the TET-TDG pathway, the exact mechanism, the targeting and the genomic regions involved remain to be investigated.

NUCLEAR RECEPTOR SUPERFAMILY

A potential site-specific, active DNA demethylation machinery must be targeted reliably to defined CpGs in regulatory sequences. In the case of hormone-induced demethylation events, a likely candidate is the hormone-activated receptor. Consistently, nuclear hormone receptors have been linked to several demethylation events [332, 468, 473-475] and found to interact with TDG [458-460] and GADD45 [476].

Nuclear hormone receptors are a class of ligand-activated proteins that translate extracellular signals into transcriptional responses. Four different classes are defined to date; steroid receptors (class I), RXR heterodimers (class II), homodimeric orphan receptors (class III) and monomeric orphan receptors (class IV). Nuclear hormone receptors consist of three main domains; a central DNA binding domain (DBD), a C-terminal ligand binding domain (LBD) and a variable transactivation domain.

The DBD of nuclear receptors consists of two highly conserved zinc fingers [477, 478]. Since they act as transcription factors, nuclear receptors rely on specific sequence motifs for binding, known as hormone response elements. Except for the monomeric orphan receptors, they bind as dimers to a sequence motif with two half sites, separated by random spacer sequences of defined length depending on the receptor. Steroid receptors typically bind to palindromic inverted repeats spaced

by 3 base pairs [479], whereas retinoic acid receptors (RAR) preferentially bind to direct repeats with a 5 base pair spacer [480].

The LBD works as a switch, changing the conformation and the properties of the receptor upon ligand binding, thereby activating the transcription factor [481, 482]. The majority of NRs has two activation function (AF) domains. The N-terminal AF1 mediates ligand-independent activation, whereas AF2 at the C-terminus is responsible for ligand-mediated effects [483]. The two AF domains cooperate to ensure full transcriptional activity of the receptor.

Steroid receptors are bound by heat shock proteins in the cytoplasm that dissociate upon hormone binding enabling the receptors to translocate in the nucleus, where they bind to DNA [484]. Most NRs are induced to dimerize upon ligand binding and then either directly recruit general transcription factors or bind co-activating accessory factors, such as members of the p160 family [485].

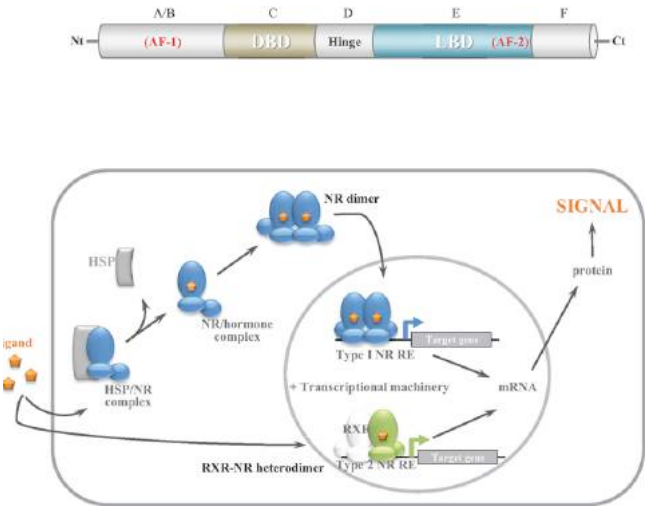


Figure 14 Structure of a typical nuclear receptor with the two activation domains (AF-1 and AF-2), the DNA binding domain (DBD) and the ligand binding domain (LBD). Below the general principle of activation of class I and class II NRs; lipophilic hormones pass the lipid bilayer of the cell and bind to their cognate receptors, causing their dissociation from heat shock proteins (HSPs) in the cytoplasm, in the case of NRs from class I. The receptors dimerize and translocated into the nucleus to exert their effector functions on target genes. In the case of NRs from class II (and III/IV), the hormones bind their receptors directly in the nucleus (class II, III, IV). From [486].

Estrogen receptors

Estrogens are involved in the regulation of gene expression in vertebrate females and, to a minor extent, in males. They are fat-soluble steroid hormones able to pass through the lipid bilayer of the cell membrane. They act through binding their cognate receptors, converting them to an active state through dissociation from the heat shock protein they are bound to in the cytoplasm. There are two estrogen receptors (ERs) in vertebrates, ERα [487] and ERβ [488], encoded by two different genes. They are highly conserved in the DBD, resulting in similar affinities towards the consensus estrogen

response element (ERE) GGTCANNNTGACC [489]. They are 59% homologous in the LBD, explaining the similar affinities to E_2 and other ligands [490]. However, ligands with distinct affinities and ligand binding characteristics for the two receptors have been described [491, 492]. One prominent example is the chemotherapeutic Tamoxifen. It acts as a selective, partial agonist/antagonist of ERα, but has a pure antagonistic effect when binding to ERβ [493]. The two ERs display different tissue distributions. ERβ is expressed primarily in thymus, testis and ovary, but absent in liver [488, 494, 495], whereas ERα is highly-expressed in liver, uterus and muscles [496].

ERα can be activated by a variety of different stimuli, but its response to E_2 is best studied. Binding of estradiol causes the receptor to translocate into the nucleus, where it exerts its effector functions by binding to palindromic EREs as a head-to-head dimer [497]. Besides the sequence motif, binding of ERα to DNA also depends on the chromatin context. Similarly to other transcription factors, it preferably binds to open chromatin regions [498]. It was shown that ERα is recruited mostly to regulatory elements distant from promoters and mediates loop formation to interact with promoters. Its depletion results in loss of chromatin loops [499]. Once bound to DNA, ERα recruits co-activators to enhance the transcriptional response. One of the first chromatin modifying enzymes found to interact with ERα was the histone acetyltransferase p300 [500] that had also been shown to interact with TDG.

ERα plays an important role in cancer development of estrogen-responsive tissues, such as mammary tissue. ERα-positive breast cancers are treatable with hormonal therapy, such as the selective estrogen receptor modulator (SERM) Tamoxifen. ERα overexpression therefore correlates with a better initial prognosis, development of resistances is frequent, however, limiting the overall prognostic value of the level of ERα in breast cancer (Reviewed in [501, 502]).

THE VITELLOGENIN SYSTEM

Vitellogenin proteins are large egg yolk precursors produced in the liver in response to estradiol. The precursor proteins are released into plasma and taken up by the oocyte, where they are cleaved into lipovitellin and phosvitin that serve as nutritional reserves in the egg. There are two minor vitellogenin genes and one major vitellogenin gene, vitellogenin II.

The chicken vitellogenin II (VTG) gene is ideally suited to study the connection between methylation, chromatin dynamics and gene activity, because it is rapidly induced upon β-estradiol treatment in terminally-differentiated hepatocytes, concomitantly with chromatin rearrangements. It is well studied and several sites in the proximal enhancer and promoter that play a role in the transcriptional control are described.

A site-specific DNA demethylation event was observed, as assessed by digests with *Hpa*II, a methylation-sensitive restriction enzyme [503]. The hormone-dependent demethylation took place in hepatocytes that express ERα and also in the oviduct that expresses ERα, but no VTG. In ERα-negative erythrocytes, there was no demethylation of the specific site. The responsible factor therefore seems to be estrogen-responsiveness in general and not transcription *per se* [503]. DNA replication was suggested not to play an important role in the demethylation, but only based on slow

cell division [504]. The next observation made was the appearance of DNaseI hypersensitive sites with different characteristics. Two sites close to the TSS were stable, while one was only transient and disappeared after hormone removal [505]. Further studies confirmed the independence from DNA synthesis by using arabinose and hydroxyurea (HU) and showed 15% undermethylation of the *Hpa*II site after 24h treatment. The same group showed in addition a relatively uniform *VTG* expression in the cells analyzed, excluding that the effect came from only a few cells within the population [506].

The DNA sequencing protocol developed by Maxam and Gilbert [507] and modified by Church and Gilbert [508] opened new possibilities to also analyze CpGs methylation status outside of *Hpa*II or other methylation-sensitive restriction enzyme recognition sites. Additionally, it was only now possible to distinguish hemimethylated sites from fully-methylated ones.

A key study in 1986 made use of this method and observed the site- and strand-specific demethylation of four CpGs in the *VTG* enhancer in liver nuclei of immature chickens. The coding (upper) strand became demethylated with kinetics parallel to those of mRNA synthesis, whereas the lower strand followed only after 24 h of hormone treatment. This re-opened the possibility that demethylation preceded transcription, which was disputed earlier because *Hpa*II cleaves DNA unmethylated on both strands. Another interesting observation concerned a CpG lying in a glucocorticoid response element. It was found to be unmethylated in the upper strand in the basal state, but 50% methylated in the lower strand. Stable presence of hemimethylated CpGs was not considered possible, because of the maintenance methyltransferase DNMT1. The same behavior of the gene in oviduct cells and erythrocytes, already described by Wilks et al., 1982, was observed here. Demethylation was taking place in liver and oviduct, expressing ERα, but not in erythrocytes. The authors made two essential conclusions from these findings: transcription of *VTG* is not needed for demethylation and binding of ERα seems to be a prerequisite for demethylation, therefore ERα is likely able to bind methylated DNA. Moreover, they revealed in preliminary results the presence of strand-specific DNA breaks in the region of the CpGs studied [332]. The follow-up study observed the binding of two non-histone proteins (NHP1 and 2) to the estrogen-responsive element (ERE) that were different from ERα and seemed to increase its binding efficiency to the ERE [509]. Another ERE was found in the *VTG* enhancer, but it has an imperfect consensus sequence and contributed only mildly to estrogen-responsiveness [510].

We decided to use this experimental system to shed light on the chromatin dynamics of the chicken *VTG* enhancer/promoter region, with a special focus on DNA methylation and the newly-discovered players in 5mC metabolism.

AIMS

It is of outmost importance for multicellular organisms to ensure the long-term stability and faithful duplication of their genomes. Simultaneously, a wide range of dynamic transactions to replicate and repair DNA, and to flexibly regulate gene expression has to be possible. Disturbances in this delicate balance readily lead to uncontrolled cell division, enhanced apoptosis or altered behavior of cells, resulting in cancer, degenerative- or autoimmune diseases, respectively. The detailed understanding of the DNA modifying processes and their regulation in cells is therefore of critical importance to prevent and treat a wide spectrum of diseases in humans.

Cellular repair processes counteracting DNA damage arising from endogenous metabolites, such as reactive oxygen species or evoked by exogenous agents, such as tobacco smoke or UV irradiation, are undoubtedly important for the survival of higher eukaryotes. This is exemplified by a variety of diseases caused by mutations in key DNA repair proteins [511, 512]. However, in addition to repairing damage arising in DNA, cells sometimes use the same pathways to generate physiological DNA modifications during development or in circumstances when the organism has to react to changes in the environment, such as infections.

This work focuses on these latter DNA modifying processes that are used by the cellular machinery for physiological purposes.

In the first part of this thesis we aimed to unravel the mechanism of class-switch recombination and the repair pathways involved. CSR is a very intriguing example of physiological genome editing that is still poorly understood. It is essential for organisms possessing adaptive immunity, but also poses a high risk, as it involves substantial DNA rearrangements. Paradoxically, several DNA repair pathways are involved in this DNA modifying process that are normally important to restore intact DNA after damage. In this interesting case, they are used to create transient damage and irreversible changes in the DNA sequence.

The second part of this thesis focuses on the DNA modification 5mC and its metabolism in DNA. 5mC has a very important role in the control of gene expression in development, as well as in somatic cells. Methylation patterns are highly deregulated in most cancers, further stressing their importance in genomic stability. To find new interventions for cancer therapy without causing side effects, it is of great importance to understand how cells maintain the sensitive balance between maintaining methylation patterns and retaining the flexibility to react to stimuli. We study the metabolism of 5mC, its modifications in DNA and its effect on transcription on the example of the chicken *VTG* gene.

RESULTS I

Published online 6 January 2016 *Nucleic Acids Research, 2016, Vol. 44, No. 6* ***2691–2705***
doi: 10.1093/nar/gkv1535

Non-canonical uracil processing in DNA gives rise to double-strand breaks and deletions: relevance to class switch recombination

Stephanie Bregenhorn[1,2], Lia Kallenberger[1], Mariela Artola-Borán[1], Javier Peña-Diaz[1,3] and Josef Jiricny[1,2,*]

[1]Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, [2]Department of Biology, Swiss Federal Institute of Technology (ETH) Winterthurerstrasse 190, CH-8057 Zurich, Switzerland and [3]University of Copenhagen, Faculty of Health Sciences Center for Healthy Aging, Department of Neuroscience and Pharmacology, Blegdamsvej 3b, DK-2200 Copenhagen N, Denmark

Received November 03, 2015; Revised December 17, 2015; Accepted December 23, 2015

ABSTRACT

During class switch recombination (CSR), antigen-stimulated B-cells rearrange their immunoglobulin constant heavy chain (C_H) loci to generate antibodies with different effector functions. CSR is initiated by activation-induced deaminase (AID), which converts cytosines in switch (S) regions, repetitive sequences flanking the C_H loci, to uracils. Although U/G mispairs arising in this way are generally efficiently repaired to C/Gs by uracil DNA glycosylase (UNG)-initiated base excision repair (BER), uracil processing in S-regions of activated B-cells occasionally gives rise to double strand breaks (DSBs), which trigger CSR. Surprisingly, genetic experiments revealed that CSR is dependent not only on AID and UNG, but also on mismatch repair (MMR). To elucidate the role of MMR in CSR, we studied the processing of uracil-containing DNA substrates in extracts of MMR-proficient and –deficient human cells, as well as in a system reconstituted from recombinant BER and MMR proteins. Here, we show that the interplay of these repair systems gives rise to DSBs *in vitro* and to genomic deletions and mutations *in vivo*, particularly in an S-region sequence. Our findings further suggest that MMR affects pathway choice in DSB repair. Given its amenability to manipulation, our system represents a powerful tool for the molecular dissection of CSR.

INTRODUCTION

Antigen-dependent antibody diversification occurs in two stages: somatic hypermutation (SHM) and class switch recombination (CSR) (1,2). During SHM, antibody-expressing B cell clones that bind antigens are stimulated to proliferate and the variable regions of their immunoglobulin (*Ig*) genes acquire a large number of mutations. Cell clones expressing mutated immunoglobulins with higher affinities for the antigen undergo positive selection and further rounds of maturation. In contrast to SHM, which yields antibodies with higher affinity through altering the variable region, CSR gives rise to antibodies of a different isotype through an irreversible rearrangement of the heavy chain (C_H) locus. The human C_H locus consists of several genes (Cμ, Cδ, $C\gamma_3$, $C\gamma_1$, $C\alpha_1$, $C\gamma_2$, $C\gamma_4$, Cϵ and $C\alpha_2$), but mature B cells express on their surface initially only IgM or IgD (the latter arising through alternative splicing). During CSR, the antigen/cytokine combination in the environment of the B cell activates transcription of specific switch (S) regions, highly-repetitive sequences flanking the 5′ termini of the C_H genes. Transcription then triggers a recombination process during which the *Ig* variable region is joined to one of the downstream constant genes with a concomitant loss of the intervening DNA. In this way, the B-cell switches from expressing the high-avidity IgM to producing high-affinity IgG, IgA or IgE with different biological effector functions (3).

SHM and CSR are initiated by activation-induced cytidine deaminase (AID), a protein expressed in antigen-activated B cells (4,5). AID converts numerous cytidines within the *Ig* locus to uracils, in a reaction that is dependent on transcription (6). In the absence of AID, both SHM and CSR are abrogated (7,8). In humans, CSR malfunction causes Hyper-IgM syndrome, characterized by elevated IgM levels and a concomitant decrease or complete absence of IgG, IgA and IgE (9).

That SHM and CSR depend not only on the generation of uracils but also on their metabolism was demonstrated by the finding that ageing mice lacking uracil-DNA N-glycosylase (UNG), an enzyme that excises uracil from

*To whom correspondence should be addressed. Email: jiricny@imcr.uzh.ch

2692 Nucleic Acids Research, 2016, Vol. 44, No. 6

DNA (10), develop B-cell lymphomas (11) and that SHM and CSR are severely attenuated in these animals (12). In humans, recessive mutations in the *UNG* gene cause Hyper-IgM syndrome (13).

Genetic evidence implicated also the mismatch repair (MMR) pathway in these processes: disruption of the mouse *MMR* genes *Msh2*, *Msh6*, *Mlh1*, *Pms2* or *Exo1* led to altered SHM and to a reduction in CSR that ranged from 2- to 7-fold, depending on the gene and the *Ig* serotype (14–21). Similarly, patients lacking PMS2 or MSH6 were diagnosed with a profound CSR defect (22,23). These findings were unexpected. Deamination of deoxycytidine, both AID-catalyzed and spontaneous (24), gives rise to U/G mispairs in DNA, however, even though these structures are recognized and bound by the human mismatch binding factor MutSα (heterodimer of MSH2 and MSH6) (25), they should not be addressed by MMR. Postreplicative MMR has evolved to remove mispaired nucleotides from the newly-synthesized strand during replication. To achieve this goal, MMR proteins need not only detect the mispair, but also direct its repair to the nascent strand. In eukaryotes, this strand is distinguished from the template by pre-existing termini, such as gaps between Okazaki fragments, where EXO1 initiates the degradation of the error-containing nascent strand up to and ∼150 nucleotides past the mispair (26). Because AID-induced U/G mispairs arise in G1 phase of the cell cycle, i.e. in DNA devoid of EXO1 loading sites, they should not trigger MMR. Instead, they should be repaired to C/G by base excision repair (BER) (27).

In all organisms, short-patch BER of uracil is initiated by the removal of this aberrant base, catalyzed primarily by UNG (10), although mammalian cells encode also the uracil-processing enzymes TDG (28), SMUG1 (29) and MBD4 (30). The resulting apyrimidinic (AP) site is then incised at its 5′ phosphate by an AP-endonuclease (APE1 in humans), which thus provides an entry site for polymerase-β (pol-β) that extends the 3′-OH terminus of the break by a single dCMP and concurrently removes the baseless sugar-phosphate residue by β-elimination. The remaining nick is then sealed by DNA Ligase III/XRCC1. Uracils can also be addressed by long-patch BER, which differs from the short-patch process in that the repair synthesis catalyzed by pol-β, pol-δ or pol-λ generates repair tracts of 2–6 nucleotides through stand displacement. This process requires, in addition to the BER enzymes, also the replication factors RFC, PCNA and FEN1 (31–33). BER-mediated repair of uracils is generally extremely efficient, possibly also due to the redundancy between UNG, TDG, SMUG1 and MBD4, however, only UNG has to date been implicated in SHM/CSR.

BER and MMR are highly-effective guardians of genomic integrity. So why are these processes linked to mutagenesis and chromosomal deletions at the *Ig* locus? One possible reason could be the high density of uracils generated by AID at its preferred target sequences WRCY (where the underlined C is the target of deamination and W = A or T; R = A or G; Y = C or T) in the regions undergoing SHM and CSR. In an earlier study, we were able to demonstrate that MMR can interfere with BER-mediated uracil repair on substrates containing a U/G and a G/T mispair in close proximity (34). We could also show that processing of substrates containing two U/G mispairs in close proximity activated PCNA monoubiquitylation and recruitment to chromatin of the error-prone DNA pol-η (35), events required for SHM (36–38). It was proposed that a similar interference between BER and MMR might give rise to the double-strand breaks (DSBs) that arise in the switch regions during the G1 phase of the cell cycle and that are necessary for CSR (39,40), but experimental evidence supporting this hypothesis is not available to date.

In an attempt to gain novel insights into the CSR process, we made use of a series of circular phagemid substrates containing uracil residues in one strand and a nick at a defined site in the other strand, the latter representing an intermediate of BER-catalyzed uracil repair and an entry point for MMR. We then studied the appearance and location of DSBs upon incubation of these molecules with MMR-proficient or -deficient extracts of human cells, as well as with purified recombinant BER and MMR proteins. We now show that DSBs can arise through processing of a combination of uracils and nicks situated on opposite strands. We also show that DSB induction is absolutely dependent on UNG, but that the dependence on MutSα is limited to substrates in which the uracils and the nick are separated by more than ∼50 nucleotides. MutLα was required only in a subset of events and this function required its endonucleolytic activity. We also provide evidence that processing of these substrates *in vivo* gives rise to deletions delineated by the uracil and/or strand break sites, and that this process is not limited to B-cells.

MATERIALS AND METHODS

Antibodies and reagents

The rabbit polyclonal anti-TDG antibody was a generous gift of Primo Schär, MSH2 antibody was from Calbiochem and the MSH6 and Lamin B antibodies were from Abcam. Restriction enzymes (NotI, XmnI, AclI and Nt.BstNBI), UDG, APE1 and the UDG inhibitor Ugi were obtained from New England Biolabs and the Protein A Dynabeads were obtained from Dynal Biotech. The purified recombinant MutLα D699N protein was a generous gift of Farid A. Kadyrov; additionally, MutLα and MutSα wt and MutSα KR (41), RPA and EXO1 were purified in our laboratory.

Cell culture

Burkitt lymphoma cells (BL2) were cultured in RPMI medium 1640 (GIBCO) with 10% fetal calf serum (GIBCO), 10 000 U/ml penicillin and 10 mg/ml streptomycin (Gibco-BRL). LoVo cells were grown in 50% DMEM (GIBCO), 50% F12 (GIBCO) supplemented with 10% fetal calf serum and penicillin-streptomycin, 293T Lα were cultured in DMEM supplemented with 10% Tet System-approved fetal calf serum (Clontech), 300 μg/ml hygromycin B (Roche) and 100 μg/ml zeocin (InvivoGen). To abrogate MutLα expression, 50 ng/ml doxycyclin (Clontech) was added to the media. siRNA transfections were carried out using the calcium phosphate transfection. Synthetic siRNA oligonucleotides sequences (all 5′ to 3′) were as follows: siLuc sequence CGTACGCGGAATACTTC-

GATT, siMSH2 UCCAGGCAUGCUUGUGUUGAATT and MSH6 CGCCATTGTTCGAGATTTA (Microsynth).

DNA substrates and DSB/mismatch repair assays

Isolation of the supercoiled homo- and heteroduplex substrates and the MMR assays were carried out as previously described (42). A heteroduplex DNA substrate (referred to as '304', Figures 1A, and 3A) containing a U/G-U/G mismatch (one within an *Acl*I restriction site) in the 89-bp polylinker of a pGEM13Zf(+) derivative pRichi was constructed by primer extension, using the primer

CCAGTGAATTGTAATA **U** GAACACTATA GGGCGAATTGGCGGCCGCGATCTGATCAGA TCCAGACGTCTGTCAA**U**GTTGGGAAGCTTGAG.

In this substrate, the distance between the nick and the closest uracil was 304 nucleotides. In order to create a substrate with a distance of 58 nucleotides between the nicking site and the closest uracil in the top strand (referred to as '58', Figure 3B), the following primer was used:

U/G-U/G 235–306 GCCCGCTTTC**U**AGTCGGGA AACCTGTCGTGCCAGCTGCATTAATG

AATCTGCCAACGCG **U** GGGGAGAGGCGG

To assess the impact of sequence context, a 95-mer oligonucleotide containing nucleotides 267–344 sequence of the human Sμ region (GenBank/EMBL/DDBJ accession no. X54713) was cloned into the PshAI site of the polylinker of plasmid pRichi. The Sμ substrate was then generated by primer extension on the phagemid ssDNA, using the oligonucleotide CCAGACGT*GGGCTAAGTTG**U**ACCAGGTGAGCTGA GCTGAGCTAGGGCTTGGCTGCACTAACTGGGCTG AG**U**TGGGCAGGGCTGGGC*CTGTCAACG. (The Sμ region sequence is shown in italics.) In this substrate, the Us are mispaired with Gs in the complementary strand.

To generate substrates with abasic sites, the following primers were used: U/G-ab: CCA GTG AAT TGT AAT A**U**G AAC ACT ATA GGG CGA ATT GGC GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCA A**X**G TTG GGA AGC TTG AG and ab-U/G: CCA GTG AAT TGT AAT **AXG** AAC ACT ATA GGG CGA ATT GGC GGC CGC GAT CTG ATC AGA TCC AGA CGT CTG TCA A**U**G TTG GGA AGC TTG AG. In all sequences, uracils are shown in bold. X represents the abasic site analog inserted with using dSpacer CE phosphoramidite (Glen Research, 10–1914–90). The desired supercoiled substrates were purified on a CsCl gradient by standard methods. The nick was introduced by incubation with Nt.BstNBI.

The DSB assays were carried out using 100 ng (47.5 fmol) heteroduplex DNA substrate and 100 μg of nuclear extracts from BL2, 293TLα+ (MLH1+), 293TLα− (MLH1−) or LoVo cells in a total volume of 25 μl in a buffer containing 20 mM Tris HCl pH 7.6, 110 mM KCl, 5 mM $MgCl_2$, 1 mM glutathione, 1.5 mM ATP, 50 μg/ml BSA. The reactions were incubated for 25 min at 30°C, followed by 60 min incubation with a stop solution (final concentrations: 0.5 mmol/l EDTA, 1.5% SDS, 2.5 mg/ml proteinase K). The DNA was purified using Qiagen MinElute Reaction Cleanup Kit followed by digestion with XmnI and treatment with 200 μg/ml RNase A (Sigma-Aldrich, Germany).

MMR efficiency was estimated as described previously (34), by digestion of the recovered DNA with AclI. In order to be able to visualize the repair tracts, the nuclear extracts were supplemented with [α-^{32}P]dATP. The DNA fragments were then separated on 1% agarose gels containing GelRed and visualized on a UV trans-illuminator. The gels were then dried and exposed to PhosphoImager screens or X-ray films.

UDG inhibition and immunodepletions

Where indicated, UDG was inhibited by the addition of Ugi peptide (4.8 units) per 100 μg nuclear extract and preincubation for 10 min at 37°C. Immunodepletions were carried out as described previously (34).

Primer extension reactions

Phagemid DNA recovered from the DSB assay was subjected to primer extension reactions, using the 5′-labeled 'Blue' GCCTCCCTCGCGCCATCAGCTTAATGCGCCGCTACAGGG or 'Red' CTCATTAGGCACCCCAGGC primers (Figure 5). The reactions were performed by adding 0.2 μM 5′-labelled primer to 50 ng DNA, 1 mM dNTPs, 1x thermobuffer and 0.3 μl Taq polymerase. The following parameters were used for the annealing and extension reactions: 15 min at 95°C, 15 min at 50°C and 35 min at 72°C. The reaction products were admixed with an equal volume of formamide and separated on 5% denaturing polyacrylamide gels. The gels were fixed in 50% EtOH, 40% H_2O and 10% acetic acid, dried and exposed to PhosphoImager screens.

Transfection/transformation experiments

HEK 293 cells were grown to 60% confluence and transfected with 1 μg U/G-U/G or C/G-C/G (ccc or nicked 5′ from the mispaired Us) plasmids (Amp resistant) and 500 ng of a control plasmid (Kan resistant) using calcium phosphate transfection. Twenty four hours post transfection, the plasmids were recovered using a DNA extraction kit (Qiagen). As DSB generation and religation was expected to be rare, we used positive selection. Half of the recovered plasmid DNA was digested with PshAI, the restriction site of which is situated in the vicinity of the uracils. Deletions were expected to eliminate this cleavage site and hence render the plasmid refractory to cleavage, while making it also unable to form colonies. DH10B bacteria were transformed with the digested or undigested plasmid DNA and plated on kanamycin or ampicillin plates. Single ampicillin-resistant colonies were picked and taken forward to colony hybridization.

Colony hybridization

The colony hybridizations were performed essentially as described (43). The colonies picked from the agar plate (see above) were streaked on a fresh plate as shown in Figure 6B. The bacteria were then transferred to a nitrocellulose membrane by placing the membrane on top of the agar plate for 30 s. The membrane was then denatured and fixed as described (43). The bacterial plate was kept at 4°C. The clones identified as mutated or deleted were then sequenced.

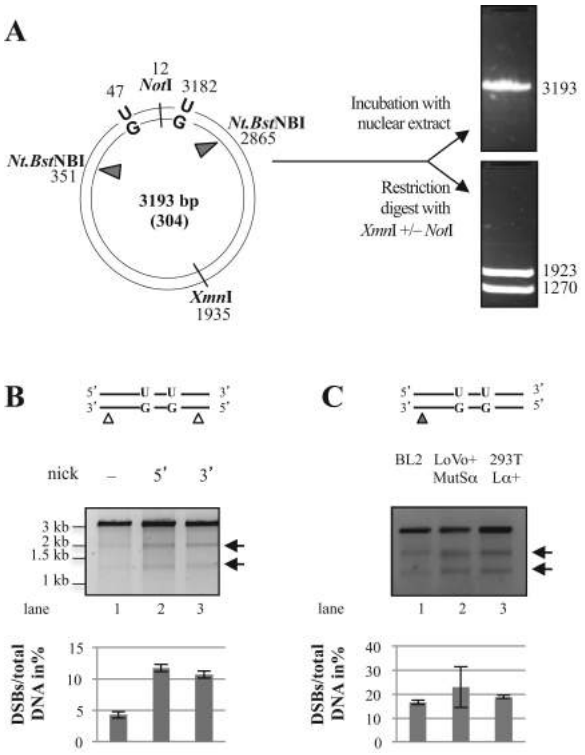


Figure 1. Processing of the U/G-U/G substrate *in vitro* gives rise to DSBs. (**A**) Schematic representation of the circular heteroduplex substrates carrying two U/G mispairs, one located in the recognition site of AclI endonuclease at nucleotide 47 and the other at position 3182. The positions of the Nt.BstNBI sites at nucleotide 351 (5′ substrate) or 2865 (3′ substrate), where a nick can be introduced selectively into the inner strand, and the position of the XmnI cleavage site are indicated. Digestion of the phagemid DNA with XmnI resulted in its linearization (3194 bp). Upon incubation of the substrates (either covalently-closed or nicked as shown in the respective figures) with the nuclear extracts, recovery and enzymatic digestion, appearance of DSBs in the vicinity of the uracils (mimicked here by additional restriction with NotI that cuts between the two uracils) gave rise to two smaller fragments of about 1900 bp and 1300 bp. The efficiency of DSB induction was estimated by ImageQuant from scans of GelRed-stained agarose gels. (**B**) Minimal requirements for DSB induction, analyzed by incubation of substrates containing two uracils in one strand and, where indicated, a nick either 3′ or 5′ from the mispaired Gs with BL2 nuclear extracts (NE). The upper panel shows UV shadowing of an agarose gel stained with GelRed. Black arrows indicate the fragments arising due to the presence of DSBs. (**C**) DSBs arising in the U/G-U/G substrate nicked 3′ from the mispaired Gs upon incubation with extracts of BL2 cells, LoVo extracts supplemented with recombinant MutSα or HEK 293TLα$^+$ (MMR-proficient) extracts. Quantifications of three independent experiments are shown in the lower panels. Error bars show mean +/− S.D. (n = 3).

Reconstituted DSB assay

For the *in vitro* DSB-assay, the reaction contained 100 ng of the substrate, 375 fmol human MutSα, 460 fmol human MutLα, 478 fmol human RPA, 26 fmol human EXOI, 1.2 fmol bacterial Uracil-DNA glycosylase and 0.378 fmol human AP Endonuclease. Briefly, the substrates were incubated with the indicated proteins in the presence of 20 mM Tris-HCl (pH 7.6), 3 mM ATP, 1 mM glutathione, 5 mM $MgCl_2$, 0.05 mg/ml BSA and 100 mM KCl. After incubation at 37°C for 20 min, the samples were heat-inactivated for 10 min at 80°C, digested with XmnI and analyzed on a 1% agarose gel containing GelRed.

RESULTS

U/G processing in B-cell extracts promotes DSB formation

AID was reported to act in a processive manner (44) and to target both strands in switch regions, albeit with different efficiencies (45). In an attempt to mimic the substrates that might arise *in vivo* in our *in vitro* assays, we set out to generate circular plasmids containing uracils in both DNA strands. However, U/G mispairs arising through AID-catalyzed deamination of cytosines are substrates for both BER and MMR and processing of substrates carrying several such mispairs could therefore potentially initiate at any uracil residue, which would complicate the analysis of the products. In order to simplify the situation, we generated substrates containing two U/G lesions in the same strand, representing the processive action of AID, and a nicking site for Nt.BstNBI on the opposite strand, either 3′

or 5′ from the uracils (Figure 1A, left panel), representing a cleaved abasic site generated upon removal of a uracil by UNG and hydrolysis by APE1 that would provide an entry site for MMR as shown in our earlier work (34). Restriction digest of these substrates with XmnI gives rise to the linearized full-length plasmid molecule of 3193 bp (Figure 1A, top right panel), but a DSB arising in the vicinity of the uracils would give rise to fragments migrating close to the 1923 and 1270 bp fragments generated by XmnI/NotI cleavage (Figure 1A, bottom right panel). When we incubated the closed-circular U/G-U/G substrates (which should be addressed by BER but not by MMR due to the absence of a pre-existing nick) with nuclear extracts (NE) of BL2 (human Burkitt lymphoma) B cells and subjected the recovered DNA to XmnI digest, only the full-length fragment was detectable (Figure 1B, lane 1). Introduction of a single nick (MMR entry site) in the opposite strand either 5′ or 3′ from the U/G mispairs prior to incubation of the substrates with the extracts yielded ∼10% of molecules cleaved by XmnI into two fragments of ∼2000 and ∼1200 bp in size, indicative of the presence of DSBs in the vicinity of the uracil residues (Figure 1B lane 1 versus lanes 2, 3).

Induction of DSBs in the nicked U/G-U/G substrate is not limited to B-cell extracts

In order to learn whether the DSBs arising as a consequence of uracil processing require B-cell specific factors, we compared the processing of the above substrates in B-cell extracts to that in extracts of LoVo and HEK293T-Lα cells. The former is a MMR-deficient human colon cancer cell line mutated in *MSH2*, but the extracts can be made MMR-proficient by complementation with purified recombinant MutSα. The latter cells are of human embryonic kidney origin. The *MLH1* MMR gene is silenced by methylation in these cells, but we have modified them to express exogenous MLH1 upon treatment with doxycycline (46). As shown in Figure 1C, DSBs were generated in nuclear extracts derived from BL2 cells as shown previously (lane 1), and even somewhat more efficiently in the MMR-proficient extracts of LoVo cells supplemented with recombinant MutSα (lane 2) or in HEK293T-Lα^{+} extracts expressing MLH1 (lane 3). This finding demonstrates that DSB generation in this system is not limited to B-cell extracts.

UNG and MutSα play key roles in DSB induction

As mentioned above, genetic evidence showed that both BER and MMR participate in CSR. We therefore wanted to test whether the induction of DSBs in our system was dependent on these repair pathways. MMR-deficient extracts of LoVo cells generated no DSBs in the U/G-U/G substrate containing a single nick in the G strand 5′ from the mispairs (Figure 2A, lane 1) and addition of the uracil DNA glycosylase inhibitor Ugi did not alter the situation (lane 2). When the MMR defect in these extracts was complemented with MutSα, DSBs were formed (lane 3), but their formation was effectively inhibited by Ugi (lane 4). This showed that the breaks were generated by a process requiring uracil excision. Similar observations were made when the substrate contained a nick in the G strand 3′ from the mispairs (Figure 2A, lanes 5–8). In order to test whether DSB induction required enzymatically-active MutSα, we supplemented the LoVo extract with the MutSα KR variant, which is mutated in the ATP binding motifs of both MSH2 and MSH6. This variant is unable to undergo the ATP-driven transition to a sliding clamp and hence remains bound at the mismatch (47). Unlike the wild type protein (Figure 2A, lane 7), MutSα KR failed to restore DSB induction in LoVo extracts (lane 9). The generation of DSBs thus requires not only the binding of MutSα to DNA, but also its ATPase activity.

We next studied the requirement for MutLα (a heterodimer of MLH1 and PMS2). When the nick was positioned 5′ from the mispaired guanines, DSBs were induced independently of MutLα (Figure 2B, lanes 1 and 3). As above, their formation was inhibited by Ugi (lanes 2 and 4). In contrast, MutLα was indispensable for the induction of DSBs in a substrate containing the nick 3′ from the mispaired guanines (Figure 2B, lane 9). No DSBs were formed when the extracts were supplemented with a MutLα variant lacking endonuclease activity (lane 7). These results conform to our understanding of MMR mechanism (see also Discussion), in which excision initiated at a 3′ nick displays an absolute requirement for MutLα, whereas excision starting at a 5′ nick does not (48). The requirement for UNG showed that DSB formation in these substrates required an intermediate of uracil processing, most likely a strand break arising through APE1-mediated cleavage of the abasic site generated by the UNG-catalyzed removal of one of the uracils. DSBs might then arise through the collision of a U/G-activated MMR-catalyzed excision tract in the nicked strand with a cleaved abasic site in the other strand.

The requirement for MMR in DSB formation is dependent on the distance between the nick and the U/G mispairs

In the *Ig* loci of activated B-cells, AID carries out multiple deaminations (44) in both strands (45). In an attempt to gain further insight into the criteria necessary to give rise to BER- and MMR-dependent DSBs *in vivo*, we varied the distance between the uracils and the nick in the opposite strand. By using uracil-containing oligonucleotides in the primer extension reaction that annealed closer to the Nt.BstNBI site, we were able to generate an additional substrate (Figure 3B), in which the nicking site was 58 nucleotides 5′ from the G of the nearest U/G mispair rather than 304 (Figure 3A). When the nick was 58 nucleotides away, the DSBs were formed with similar efficiency as in the substrate where it was 304 nucleotides away (Figure 3C). [NB: In the '58' substrate, a DSB generated close to the uracils and XmnI cleavage generate two fragments of similar size, hence only a single band is seen in the agarose gel, as opposed to two different size bands generated from the '304' substrate.]

Because CSR efficiency is only decreased and not abolished in MMR-deficient cells, we wanted to test whether the DSBs generated in the '58' substrate were MMR-dependent, like those detected in the '304' substrate (Figure 2). As shown in Figure 3D, incubation of the former substrate with MMR-deficient extracts of LoVo cells re-

2696 Nucleic Acids Research, 2016, Vol. 44, No. 6

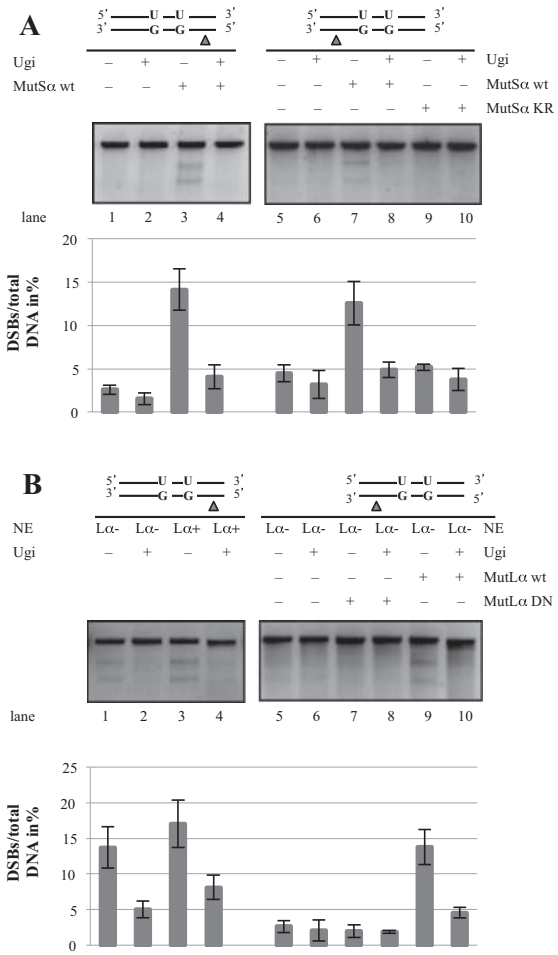


Figure 2. UNG and MutSα play key roles in DSB induction. (**A**) Lanes 1–4: DSB formation in a U/G-U/G substrate, nicked 5′ from the mispaired Gs, upon incubation with NE of LoVo cells supplemented with recombinant MutSα and pretreated with Ugi where indicated. Lanes 5–10: Incubation of the U/G-U/G substrate, nicked 3′ from the mispaired Gs, with LoVo extracts supplemented either with wild type MutSα or the ATPase-dead KR mutant, and pretreated with Ugi where indicated. (**B**) Lanes 1–4: DSB formation in a U/G-U/G substrate, nicked 5′ from the mispaired Gs, upon incubation with MMR-proficient (HEK 293T-Lα^{+}) and -deficient (HEK 293T-Lα-) extracts. UNG was inhibited by the addition of Ugi where indicated. Lanes 5–10: DSB formation in a U/G-U/G substrate, nicked 3′ from the mispaired Gs, upon incubation with MMR-proficient (HEK 293T-Lα^{+}) or -deficient (HEK 293T-Lα^{-}) extracts. The MMR-deficient extracts were supplemented with either wild type MutLα or its endonuclease-deficient mutant (DN). Extracts were pretreated with Ugi where indicated. Quantifications of three independent experiments are shown in the lower panels. Error bars show mean +/− S.D. (n = 3).

sulted in the generation of DSBs when the extract was supplemented with purified recombinant MutSα (lane 3), but were detectable, albeit only weakly, even in the unsupplemented extract (lane 1). Because DSB formation was still absolutely dependent on UNG (lanes 2,4), we postulated that they must have arisen by MMR-independent exonucleolytic degradation between the nick and the cleaved abasic site(s) in the opposite strand. This was substantiated by the finding that no DSBs were detected upon incubation of this substrate with extracts of EXO1-deficient cells (Supplementary Figure S1).

The DSB formation is dependent on sequence context

As mentioned in the Introduction, deletions occurring at the *Ig* locus during CSR initiate and terminate at the intergenic S regions. In order to learn whether these repetitive sequence elements affect the efficiency of DSB formation, we cloned a fragment of the Sμ-region into our

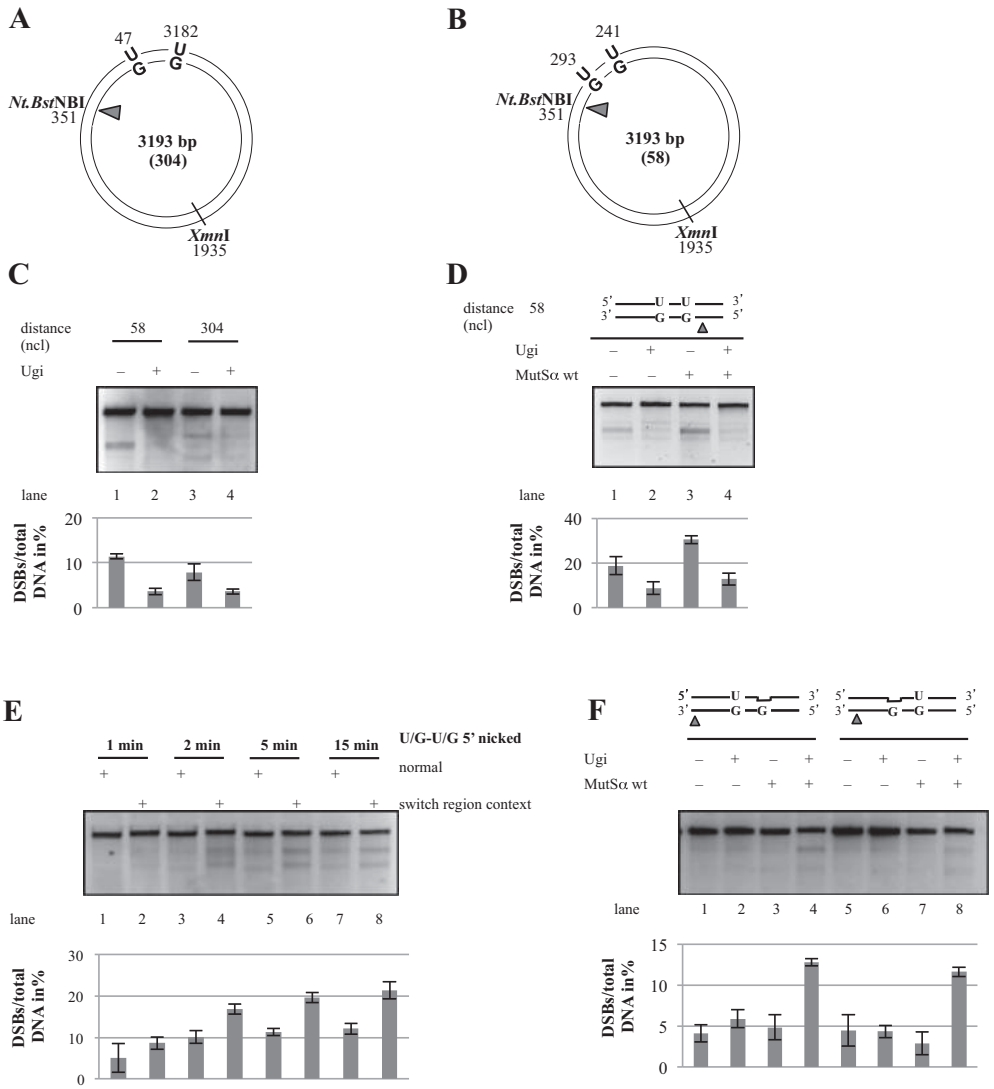


Figure 3. Lesion density and sequence context affect DSB induction. Schematic representation of the circular heteroduplex substrates 304 (**A**) and 58 (**B**) carrying two U/G mispairs and a Nt.BstNBI site 58 or 304 nucleotides 5′ from the nearest mispaired G. (**C**) The Nt.BstNBI-nicked '58' or '304' substrates were incubated with BL2 extracts pretreated with Ugi where indicated. (**D**) The Nt.BstNBI-nicked '58' substrate was incubated with LoVo extracts pretreated where indicated with Ugi and supplemented or not with recombinant MutSα. (**E**) The presence of uracils and a nick located within a sequence fragment from an *Ig* switch region renders the phagemid more prone to DSB induction in a time course experiment. (**F**) Same as in C, but the experiment was performed with '304' substrates containing an abasic site, a U/G mispair and a nick in the G-strand 3′ from the U/G mispairs as indicated. Bar graphs show quantifications of three independent experiments. Error bars show mean +/− S.D. (n = 3).

phagemid and positioned the U/G mispairs 52 nucleotides apart within this sequence (see Materials and Methods). As shown in Figure 3E, DSB induction in the switch region sequence context was more efficient than in the original phagemid (compare lanes 1,3,5,7 with lanes 2,4,6,8). This result suggested that the S regions might be highly recombinogenic *in vivo* not only because they contain numerous AID consensus sequences, but possibly also because the highly-repetitive sequences might form secondary structures upon degradation of one strand. This would hinder repair synthesis and thus increase the time window for a collision of the MMR and or BER degradation tracts, which could result in DSB formation.

MMR involvement in DSB generation requires U/G mispairs

Exonucleolytic degradation of the error-containing strand during postreplicative MMR is initiated by loading of EXO1 at the nick by mismatch-activated MutSα. We wanted to confirm that the MMR-dependent DSB formation observed in our assays was indeed triggered by MutSα binding to the U/G mispairs. We also wondered whether the low efficiency of DSB formation was linked to the rapid removal of the uracils by UNG, which would convert our U/G-U/G substrate into a DNA molecule carrying two abasic sites, which are not substrates for MutSα. We therefore generated a substrate carrying only a single U/G mispair and a synthetic abasic site (a tetrahydrofuran ring) in place of the second uracil. This substrate mimics the U/G-U/G substrate from which one uracil had been removed by UNG. When a nick was introduced 3′ or 5′ from the mispair into the unmodified G strand, DSBs arose only in the MMR-proficient extract and upon inhibition of UNG with Ugi (Figure 3F, lanes 4,8). This suggested that UNG-catalysed removal of the uracil was more rapid than recognition of the U/G mispair by MutSα. Once the uracil was removed, the substrate contained either two abasic sites (intact or cleaved), or two C/G pairs generated by successful BER (see next section). As neither abasic sites nor C/G pairs are substrates for MMR, the strand degradation process was not activated.

U/G mispairs are processed by BER more rapidly than by MMR

We could test the above hypothesis directly. The 3′ uracil in the U/G-U/G substrate is situated in an AclI site such that it renders the substrate refractory to cleavage by this enzyme (Figure 4A). Correction to C/G, which can be mediated either by BER or by MMR directed to the uracil-containing strand, restores the AclI site (35). Because the repaired plasmid contains three AclI sites as opposed to only two in the U/G-U/G substrate, digestion of the repaired plasmid with AclI gives rise to three fragments of 1515, 1305 and 373 bp, whereas the unrepaired plasmid is cleaved into only two, of 2820 and 373 bp. As shown in Figure 4B (lane 1), incubation of the U/G-U/G plasmid nicked in the G strand 5′ from the mismatch in the AclI site with MMR-proficient extracts of HEK293TLα$^+$ cells resulted in ∼60% repair to C/G. Importantly, much of this repair was mediated by BER, as its efficiency was reduced by maximally 50% (lanes 2,3) in the

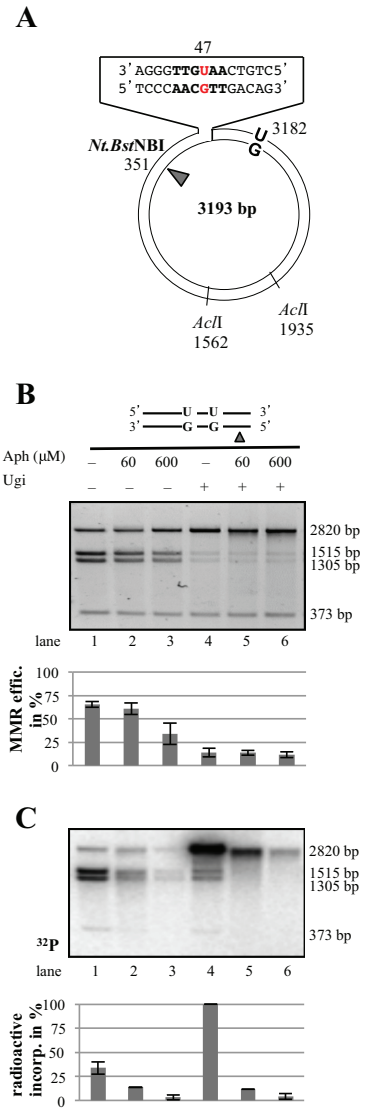


Figure 4. Both BER and MMR are activated on the nicked U/G-U/G substrate. (**A**) Schematic representation of the circular heteroduplex MMR substrate carrying two U/G mispairs, one located in the recognition site of AclI endonuclease at nucleotide 47 and the other at position 3182. The positions of the Nt.BstNBI site, where a nick can be introduced selectively into the inner strand 5′ from the mispaired Gs, and the position of the other two AclI cleavage sites are indicated. (**B**) U/G to C/G mismatch repair efficiency estimated by conversion of AclI-resistant to AclI-susceptible substrate (34). Aph, aphidicolin. The panel shows a representative image of a 0.8% agarose gel stained with GelRed. (**C**) Autoradiograph of the dried 0.8% agarose gel shown in panel B above. Bar graphs in panels B and C show quantifications of three independent experiments. Error bars show mean +/− S.D. (n = 3).

presence of aphidicolin, an inhibitor of DNA polymerase δ that carries out repair synthesis in the *in vitro* MMR reaction (26). Moreover, the MMR-mediated events were directed into the U-containing strand, most likely through repair tracts initiating at the cleaved abasic site generated by BER at the 5′ uracil (34), because only U/G to C/G repair restores the AclI site. This was further confirmed by the incorporation of [α-^{32}P]dAMP into the 1515 and 1305 bp AclI fragments that only arise upon U/G to C/G repair (Figure 4C, lane 1).

In contrast, inhibition of UNG with Ugi inhibited AclI cleavage almost completely (Figure 4B, lanes 4–6). Importantly, MMR was activated on this substrate more efficiently than in the absence of Ugi, as witnessed by the greater incorporation of [α-^{32}P]dAMP into the phagemid molecules (Figure 4C, lane 4). Moreover, this process was mostly directed to the G-strand, given that the labeled molecules were refractory to AclI cleavage. This is indicative of U/G-activated MMR repair tracts initiating at the Nt.BstNBI nicking site, which would convert the U/G mispairs into U/A pairs.

Taken together, the latter evidence helps explain why DSB incidence in our system is relatively low. First, to generate a DSB, the MMR repair tract must collide with a BER intermediate in the opposite strand. Second, in order for MMR to be activated, some U/G mispairs must remain in the DNA, which is likely to occur only rarely due to the high efficiency of BER.

In vitro-generated DSBs cluster mostly in the proximity of the uracil sites

Using primer extension assays, we mapped the position of the DSBs arising in the U/G-U/G substrate nicked 3′ from the mispairs in the G-strand upon incubation with extracts of MMR-proficient HEK 293TLα^{+} cells. The most prominent breakpoint in the G-strand (mapped with the 'blue' primer) was located in close proximity of the AclI site containing the distal (3′) uracil residue (Figure 5A, lane 1). Upon inhibition of UNG with Ugi (lane 2), several weaker bands were detected, which were most likely generated by the MutLα heterodimer (49) upon MMR activation by the U/G mismatch.

The major breakpoint in the uracil-containing strand, mapped by extension of the 'red' primer, also resided at or close to the uracil within the AclI site (Figure 5B, lane 1). As addition of Ugi to the extract substantially reduced the incidence of this breakpoint (lane 2), we propose that it arose through APE1-catalyzed cleavage of the abasic site generated by UNG. Under the latter conditions, a second, slightly shorter, product was also detected. This fragment might represent an intermediate of uracil processing by long-patch BER initiated by other glycosylases present in the extracts (MBD4, SMUG1 or undepleted TDG).

DSBs in the U/G-U/G substrate are induced also *in vivo*

The sizes of the plasmid fragments generated from the U/G-U/G substrates in our *in vitro* assay indicated that the DSBs were occurring in close proximity to the uracils. We wanted to learn whether these findings reflected the processing of these substrates *in vivo*. DSBs arising upon transfection of the U/G-U/G plasmids into human cells should be repaired by non-homologous end-joining (NHEJ), a process that generally leads to a small loss of genetic information at the break (50,51). We therefore designed a protocol that enabled us to identify the mutated plasmids among the large excess of molecules that were corrected by BER to C/G-C/G without mutations. We carried out these experiments in human kidney 293 (HEK 293) cells, because they can be transfected with very high efficiency. Moreover, we showed above that DSBs form in uracil-containing substrates also in non-B cell types, and ectopic expression of AID in non-B cells was shown to trigger both SHM and CSR (52–54), suggesting that AID is the only B-cell-specific factor required for these processes.

We transfected the U/G-U/G or the control C/G-C/G substrate into the HEK293 cells, either wild-type or depleted of MutSα with siRNAs targeting MSH2 and MSH6 (Figure 6A), recovered the plasmid DNA 24 h later and used it to transform competent *E. coli* after digestion with PshAI. The cleavage site of this enzyme is situated in close proximity of the uracils and it was therefore anticipated that all plasmid molecules carrying the wild-type sequence at this site would be linearized and thus made transformation-incompetent. [NB: The enrichment for mutant molecules

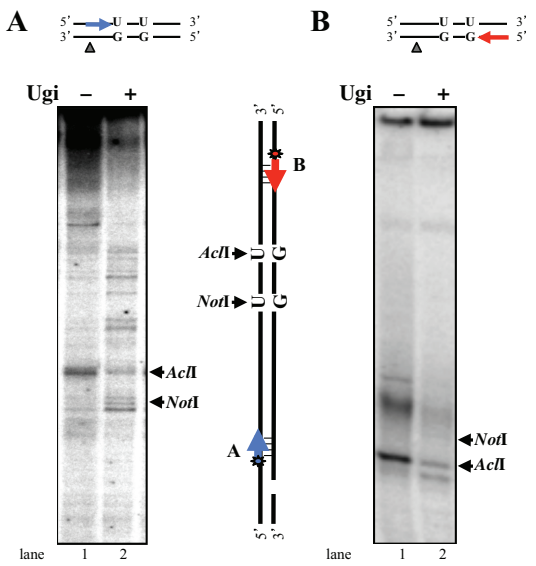


Figure 5. Breakpoints in the U/G-U/G substrate are at or near the uracil residues. U/G-U/G substrate nicked 5′ from the mispaired Gs was incubated with HEK 293T-Lα$^{+/-}$ extracts pretreated with Ugi where indicated. The DNA was recovered and subjected to primer extension using 5′ labeled primers 'blue' and 'red' (see Materials and Methods). The reaction products were separated on 5% denaturing polyacrylamide gels. Markers were products of primer extensions on plasmids digested with *Acl*I (A) or *Not*I (N). The 'blue' primer was complementary to the G strand (**A**), the 'red' primer was complementary to the uracil-bearing strand (**B**) as shown schematically between the panels. The figure shows representative autoradiographs from four independent experiments.

2700 Nucleic Acids Research, 2016, Vol. 44, No. 6

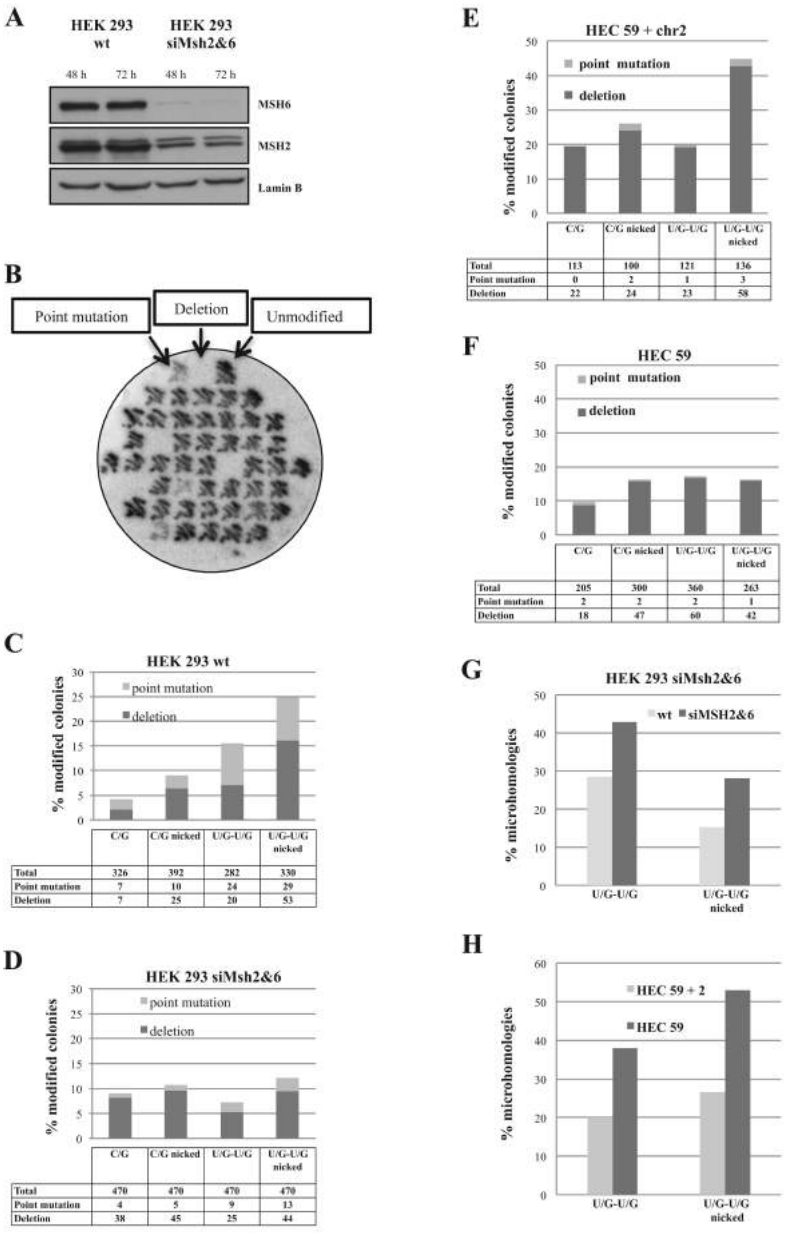


Figure 6. *In vivo* repair of DSBs in the U/G-U/G substrate leads to mutations and deletions. (**A**) Western blot of extracts of HEK 293 cells showing efficiency of knock-down of MSH2 and MSH6 with siRNA. (**B**) Autoradiograph of a representative filter carrying plasmid DNA recovered from transfections of HEK 293 cells with the U/G-U/G substrate nicked 5′ from the mispaired Gs, hybridized with a radiolabeled oligonucleotide complementary to the wild type (unmodified) plasmid sequence flanking the uracil residues. (**C**) Mutated and deleted plasmids recovered from multiple transfections of HEK 293 cells with the listed substrates. (**D**) As in C, but the substrates were transfected into HEK 293 cells depleted of MutSα. (**E**) As in C, but the substrates were transfected into HEC 59 + chr2 cells (MMR-proficient). (**F**) As in C, but the substrates were transfected into HEC 59 cells. (**G**) Incidence of microhomologies in plasmids recovered from transfections of the indicated substrates into HEK 293 cells depleted or not of MutSα. (**H**) Incidence of microhomologies in plasmids recovered from transfections of the indicated substrates into HEC 59 + chr2 or HEC 59 cells.

by digestion with PshAI was necessary, because the formation of DSBs and/or uracil misrepair *in vivo* are extremely rare.] In six independent experiments, we generally obtained ∼20 times more colonies from transformations with the recovered PshAI-digested U/G-U/G plasmid than with the C/G-C/G homoduplex control (data not shown). However, DNA sequencing revealed that the majority of these molecules in both populations were devoid of mutations, which implied that the PshAI digest was not quantitative. We therefore introduced a second screening step for the mutants: colony hybridization with a radiolabeled oligonucleotide complementary to the sequence flanking the PshAI site. We then sequenced only those colonies that indicated the presence of point mutations or deletions (Figure 6B).

As shown in Figure 6C and Supplementary Figure S2A, of 326 sequenced colonies from control transformations of HEK 293 cells with the covalently-closed circular (ccc) C/G-C/G plasmid, 7 (2.1%) contained point mutations and 7 (2.1%) contained deletions. Because the mutants were selected with a probe located between the two uracil residues, all examined deletion mutants lost this site. DNA sequencing revealed that these very rare deletions in the unnicked C/G-C/G substrate were several hundred base pairs long and did not share their end points. The presence of a nick in the G strand of this substrate trebled the occurrence of these large deletions to 6.3% and, as anticipated, one end point of these deletions was often (∼30%) close to the position of the nick. Transfection of the ccc U/G-U/G substrate yielded both mutations (8.5%) and deletions (7%), which were generally short, mostly <200 bp. Transfection of the nicked U/G-U/G substrate yielded plasmids containing a similar amount of mutations (8.7%), but substantially more deletions (16%), largely spanning the distance between the position of the uracils and the Nt.BstNBI nick. Most of the end-joining events occurred at sequences with no apparent homologies. This resembled analysis of hybrid S-S junctions in immunoglobulin switch regions, which showed that the sequences were fused at or very close to AID hotspots (40). We also detected several translocation events between the U/G-U/G plasmid substrate and genomic DNA (Supplementary Figure S2A, yellow lines). This finding shows that DSBs arising through the interplay of MMR and BER during the processing of proximal uracil residues in our substrate are recombinogenic and provides independent support for the hypothesis that chromosomal translocations arising in activated B cells (55) might have arisen through a similar mechanism.

We next wanted to study the effect of the MMR system on the above events. To this end, we treated the HEK 293 cells with siRNA targeting MSH2 and MSH6 to deplete MutSα (Figure 6A). Transfections of the above substrates into these cells yielded some unanticipated results (Figure 6D, Supplementary Figure S2B), inasmuch as the proportion of deletions was similar in all four substrates (8% for ccc C/G-C/G, 9.5% for nicked C/G-C/G, 5.3% for ccc U/G-U/G and 9.3% for nicked U/G-U/G), but—most remarkably—the number of point mutants was substantially reduced (0.8% for ccc C/G-C/G, 1% for nicked C/G-C/G, 1.9% for ccc U/G-U/G and 2.7% for nicked U/G-U/G). In all these mutants, the deletions were large and the influence of the nick was much less pronounced, which confirms that the nick is a specific signal for MMR-catalyzed strand degradation events.

We then transfected these substrates into the MMR-deficient human endometrial cancer cell line HEC 59, which is mutated in *MSH2*, or HEC 59 + chr2 cells, in which the MMR defect was corrected by transfer of chromosome 2 that carries the wild type *MSH2* gene. Unexpectedly, the number of point mutants identified in both cell lines was very low, but the number of deletions in the MMR-proficient cells (Figure 6E, Supplementary FIgure S2C) was more than 2.5-fold higher than in the MMR-deficient line (Figure 6F, Supplementary FIgure S2D), particularly in the nicked U/G-U/G substrate. Their size and end-points reflected those observed in HEK 293 cells. Interestingly, sequence analysis of the clones revealed an increase in the number of microhomologies around the break points in the plasmids recovered from MMR-deficient MutSα-depleted HEK 293 (Figure 6G) and HEC 59 cells (Figure 6H), which suggests that the molecular mechanism of DSB repair is affected by the MMR status of the cells. [N.B.: The sequencing data are not shown due to space limitations, but are available upon request.]

DSBs arise in the nicked U/G-UG substrate in a system reconstituted from purified proteins

Based on the results of the above experiments and on published biochemical (34,35) and genetic (40,56,57) data, we postulated that the DSBs in the nicked U/G-U/G substrate might arise by a mechanism outlined in Figure 7. To substantiate that the DSBs in our system arose by this mechanism and that the process did not involve additional, as yet unidentified, factor(s) present in human cell extracts, we reconstituted the repair process from purified recombinant UNG, APE1, MutSα, MutLα, RPA (required to stabilize the single-stranded DNA gap) and EXO1. As shown in Figure 8, DSB generation in the substrate nicked in the G-strand 5′ from the mispairs was rather efficient.

DISCUSSION

Antigen-induced B cell activation leads to the expression of AID, which is targeted to the *Ig* loci, where it deaminates a subset of cytosines in its preferred recognition sequences to uracils. Metabolism of the AID-generated uracils then triggers SHM, during which the rearranged VDJ regions undergo extensive mutagenesis, and CSR that affects the switch regions between the constant genes. Here, uracil processing results in the generation of not only mutations, but also DSBs that trigger non-homologous end-joining and cause an irreversible loss of the intervening DNA sequences. Genetic evidence implicated BER and, to a lesser extent, also MMR in both these processes, even though the absence of UNG resulted in a substantially more severe phenotype than MMR deficiency. However, the absence of both MMR and BER abolishes CSR completely (58,59). What is the nature of the molecular transactions that allow SHM and CSR to subvert antimutator DNA repair processes such as BER and MMR to give rise to mutations and genomic rearrangements?

2702 Nucleic Acids Research, 2016, Vol. 44, No. 6

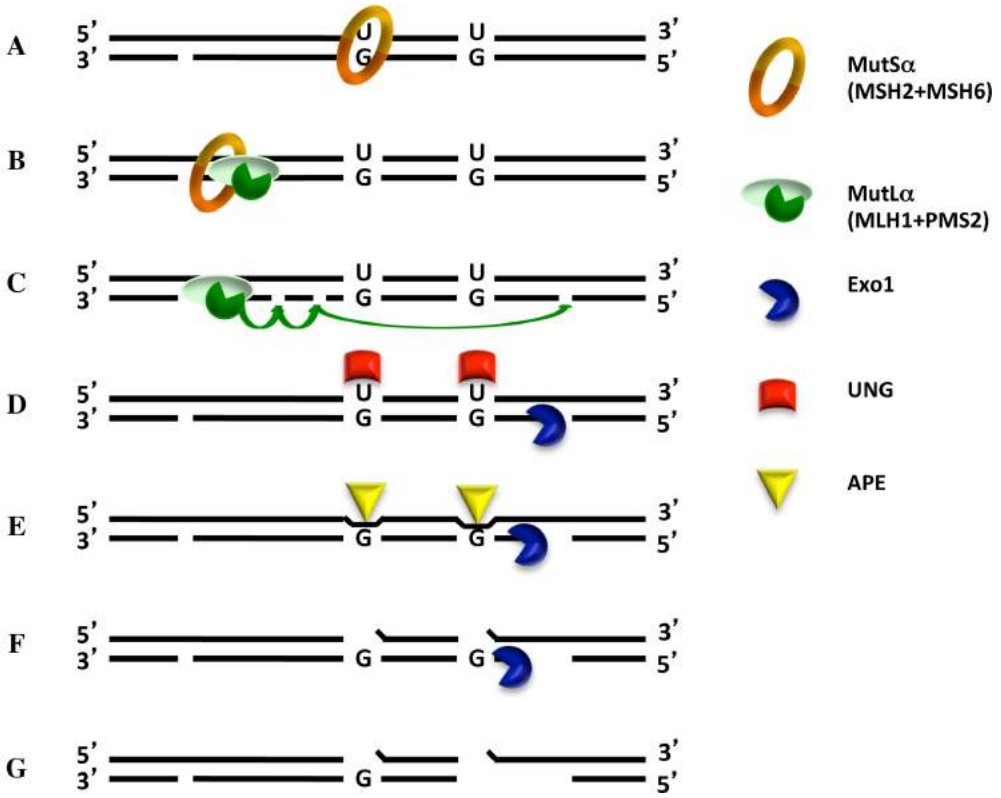


Figure 7. Putative scheme of BER- and MMR-dependent DSB induction in *Ig* loci. The first step of CSR is the active recruitment to the *Ig* locus of AID, which deaminates several cytosines in both DNA strands. The U/G mispairs generated at these sites are substrates for both BER and MMR. In a scenario explored in the present study, removal of a uracil in the bottom strand by UNG followed by APE1-mediated cleavage of the abasic site would generate the substrate shown in (**A**). Recognition of one of the remaining U/G mispairs by MutSα/MutLα causes activation of the complex and its translocation along the DNA contour. Interaction with PCNA loaded at the cleaved abasic site (**B**) activates the endonuclease activity of MutLα, which, upon complex formation with PCNA, introduces additional nicks into the discontinuous (bottom) strand (**C**). One of these is used as an entry site for EXO1, which commences to degrade the DNA in a 5′ to 3′ direction (**D**). If the remaining Us in the upper strand were in the meantime addressed by UNG (**E**), APE1 will cleave the abasic sites (**F**). Collision of the EXO1-catalyzed excision tract in the lower strand with the APE1-induced break in the upper strand will give rise to a DSB and EXO1 dissociation (**G**). [N.B. In this scenario, DSB formation is absolutely dependent on MutLα, which has to introduce additional nicks into the bottom strand, some of which must be situated 5′ from the mispaired Gs. Only then will EXO1 be able to degrade this strand toward the U/G mispairs. Had the process been initiated by the removal of a uracil residue 5′ from the two mispaired Gs, the process would have been largely independent of MutLα.]

We argued that one answer to this question might lie in the processive action of AID. Because this enzyme is targeted to the variable and switch regions of the *Ig* locus (60), uracils arise at these sites frequently and are clustered together, rather than being dispersed and rare, as is the case when they arise through spontaneous hydrolytic deamination. This unique situation provides both BER and MMR with substrates in the same region of DNA and, because the two repair systems do not normally encounter each other, it is likely that there are no evolutionary safeguards in place that prevent them from interfering with one another. Our experiments suggest that the incidence of DSBs arising through BER and MMR interference is rather low, even in our system that was designed to increase the likelihood of their detection. *In vivo*, the mutagenic events triggering SHM/CSR also occur rather infrequently, but they are detectable due to a process of positive selection of the mutated and/or rearranged clones. Moreover, CSR can be triggered by only a few breaks, providing that they occur at sites conducive to recombination, such as the *Ig* switch regions. Experimental evidence in support of the hypothesis that strategically-placed DSBs are sufficient for CSR was provided by replacement of the switch regions with yeast I-*Sce*I endonuclease cleavage sites (61).

The finding that CSR-associated DSBs arise outside of S-phase (40) rules out the possibility that these DSBs are generated when replication forks encounter single-strand breaks arising through APE1-mediated cleavage of abasic sites left behind after uracil removal. In this study, we provide evidence supporting this hypothesis, inasmuch as the

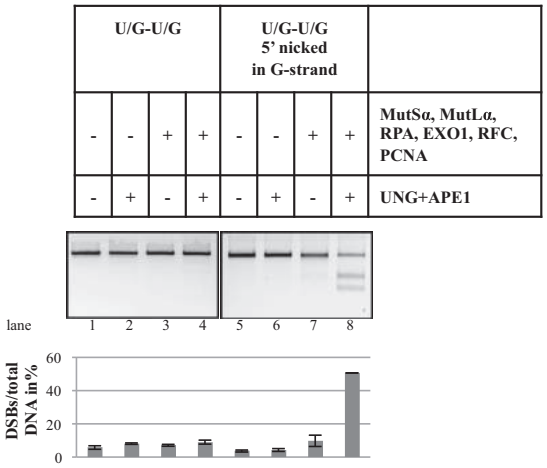


Figure 8. Reconstitution of the minimal BER/MMR interference system. The minimal mismatch repairosome was assembled from purified recombinant MutSα, MutLα, RPA, EXO1, RFC and PCNA. Uracil processing was mediated by purified recombinant UNG and APE1. Following incubation of the U/G-U/G and nicked U/G-U/G substrates with the indicated protein combinations, the recovered plasmids were subjected to electrophoresis on 0.8% agarose gels. The panel shows an image of a representative gel stained with GelRed. Bar graph shows quantification of three independent experiments. Error bars show mean +/− S.D. (n = 3).

DSBs arising in our *in vitro* assay arise in DNA molecules that do not replicate and that therefore lack strand discrimination signals (such as Okazaki fragment termini) that MMR uses to direct excision to the nascent DNA strand. Indeed, no DSBs were generated in the covalently-closed U/G-U/G substrate even when BER was active (Figure 1), nor were they formed in the same substrate when BER was blocked by the addition of Ugi, even though a nick was present in the G-strand (Figures 2–4). In contrast, when BER was active, DSBs were readily detected, providing that the substrate was nicked and that MMR was also active. This shows that the breaks in our system were generated by interplay of the two pathways, where MMR-dependent strand excision initiated at the nick in one strand collided with a cleaved abasic site arising after uracil removal and in the opposite strand. This hypothesis is experimentally substantiated by our break- and deletion-mapping experiments, which showed that the DSBs were predominantly positioned at or close to one of the uracils (Figure 5) and that the deletions most often spanned the distance between the nick and the closest uracil (Supplementary Figure S2). It should be remembered that the nick in our model substrates acts as a surrogate for an APE1-cleaved abasic site arising through UNG-catalyzed uracil removal, which we could show previously to act as a loading site for MMR-activated EXO1 (34).

During processive AID-catalyzed cytosine deamination of the *Ig* loci, uracil residues would be generated in both strands in close proximity. Although BER is normally extremely efficient, uracil repair in antigen-activated B cells appears to be sluggish. This leads on the one hand to mutations caused probably to Rev1/pol-ζ-catalyzed error-prone by-pass of a subset of abasic sites arising at the sites of deamination upon uracil removal by UNG and, on the other hand, to strand breaks generated by APE1-catalyzed cleavage of another subset of these abasic sites. Thus, AID-deaminated S region might contain intact U/G and G/U mispairs, as well as incised and intact abasic sites. The intact mispairs would activate MMR, which would load EXO1 at the strand discontinuities to generate a single-stranded excision tract that could be up to a kilobase long. If this tract were to collide with a cleaved abasic site or indeed with a MMR-activated EXO1-catalyzed excision tract in the opposite strand, a DSB would arise.

Our transfection experiments yielded also an unexpected finding, in that the frequency of microhomologies was higher in MMR-deficient cells compared to the MMR-proficient controls (Figure 6). The requirement for MMR in the control of HR fidelity has long been known (62), but its involvement in NHEJ or microhomology-mediated end-joining (MMEJ) has to our knowledge not been reported to date in recombination-proficient cells.

DSBs are a frequent cause of chromosomal rearrangements that play a key role in cancer, particularly in instances where recombination is disrupted, such as in case of BRCA1/2 mutations that cause breast cancer predisposition (63). In B-cells, deregulation of AID expression and its aberrant targeting induces translocations and results in lymphomagenesis (64), and it would appear highly likely that the latter chromosomal rearrangements are triggered by DSBs arising through MMR- and BER-mediated processing of AID-generated uracils. As shown above, the interference of MMR and BER is neither specific to B-cells, nor to U/G processing. We therefore posit that this phenomenon could come into play in instances where DNA modifications, such as U/G mispairs, addressed by both systems are generated at high density. One analogous example is methylation damage induced by S_N1 type methylating agents, such as the cancer therapeutics temozolomide, dacarbazine or streptozotocin, which generate substrates for both BER and MMR and which are highly toxic to MMR-proficient, but not –deficient cells (46). It is also possible that the interference of MMR and BER may be the cause of the MMR-dependent DNA damage-induced chromosomal instability in colon cancer cells (65). Indeed, MMR-deficient colon cancer cell lines are generally diploid, in contrast to MMR-proficient cells, which are frequently aneuploid (8).

DNA repair processes are generally thought of as guardians of our genome. In this work we demonstrate that their canonical roles can—under certain circumstances—cause genomic instability. In this particular case, mutagenesis and genetic deletions that accompany SHM/CSR are beneficial to the organism, given that they protect it from infections. Moreover, these processes affect B cells that are programmed to die, which minimizes the possibility of deleterious side effects. Thus, even though aberrant recombination linked to antibody diversification may in rare instances cause cancer, the benefits offered by SHM/CSR in evolutionary terms far outweigh the drawbacks. It is our hope that an understanding of the molecular mechanisms of the latter processes will help us

find the Achilles heel of cells in which antibody diversification went awry and, in the long run, identify ways of eliminating them.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Anne Durandy for helpful advice and numerous stimulating discussions, to Anja Saxer and Kalpana Surendranath for expressing and purifying the recombinant MutSα and MutLα, and to Stefano Ferrari and Pavel Janscak for the gift of the EXO1 and RPA expression vectors, respectively.

FUNDING

Swiss National Science Foundation [310030B-133123 and 31003A-149989]; European Research Council ['MIRIAM' 294537 to J.J.]. Funding for open access charge: University of Zurich.
Conflict of interest statement. None declared.

REFERENCES

1. Vaidyanathan,B., Yen,W.F., Pucella,J.N. and Chaudhuri,J. (2014) AIDing chromatin and transcription-coupled orchestration of immunoglobulin class-switch recombination. *Front. Immunol.*, **5**, 120.
2. Saribasak,H. and Gearhart,P.J. (2012) Does DNA repair occur during somatic hypermutation? *Semin. Immunol.*, **24**, 287–292.
3. Manis,J.P., Tian,M. and Alt,F.W. (2002) Mechanism and control of class-switch recombination. *Trends Immunol.*, **23**, 31–39.
4. Muramatsu,M., Kinoshita,K., Fagarasan,S., Yamada,S., Shinkai,Y. and Honjo,T. (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, **102**, 553–563.
5. Muramatsu,M., Sankaranand,V.S., Anant,S., Sugai,M., Kinoshita,K., Davidson,N.O. and Honjo,T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.*, **274**, 18470–18476.
6. Chaudhuri,J. and Alt,F.W. (2004) Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.*, **4**, 541–552.
7. Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Labelouse,R., Gennery,A. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, **102**, 565–575.
8. Lengauer,C., Kinzler,K.W. and Vogelstein,B. (1997) Genetic instability in colorectal cancers. *Nature*, **386**, 623–627.
9. Durandy,A., Revy,P., Imai,K. and Fischer,A. (2005) Hyper-immunoglobulin M syndromes caused by intrinsic B-lymphocyte defects. *Immunol. Rev.*, **203**, 67–79.
10. Lindahl,T. (1974) An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 3649–3653.
11. Nilsen,H., Stamp,G., Andersen,S., Hrivnak,G., Krokan,H.E., Lindahl,T. and Barnes,D.E. (2003) Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene*, **22**, 5381–5386.
12. Rada,C., Williams,G.T., Nilsen,H., Barnes,D.E., Lindahl,T. and Neuberger,M.S. (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.*, **12**, 1748–1755.
13. Imai,K., Slupphaug,G., Lee,W.I., Revy,P., Nonoyama,S., Catalan,N., Yel,L., Forveille,M., Kavli,B., Krokan,H.E. *et al.* (2003) Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat. Immunol.*, **4**, 1023–1028.
14. van Oers,J.M., Roa,S., Werling,U., Liu,Y., Genschel,J., Hou,H. Jr, Sellers,R.S., Modrich,P., Scharff,M.D. and Edelmann,W. (2010) PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 13384–13389.
15. Martomo,S.A., Yang,W.W. and Gearhart,P.J. (2004) A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J. Exp. Med.*, **200**, 61–68.
16. Rada,C., Ehrenstein,M.R., Neuberger,M.S. and Milstein,C. (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*, **9**, 135–141.
17. Neuberger,M.S., Lanoue,A., Ehrenstein,M.R., Batista,F.D., Sale,J.E. and Williams,G.T. (1999) Antibody diversification and selection in the mature B-cell compartment. *Cold Spring Harb. Symp. Quant. Biol.*, **64**, 211–216.
18. Ehrenstein,M.R. and Neuberger,M.S. (1999) Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *EMBO J.*, **18**, 3484–3490.
19. Schrader,C.E., Edelmann,W., Kucherlapati,R. and Stavnezer,J. (1999) Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *J. Exp. Med.*, **190**, 323–330.
20. Li,Z., Scherer,S.J., Ronai,D., Iglesias-Ussel,M.D., Peled,J.U., Bardwell,P.D., Zhuang,M., Lee,K., Martin,A., Edelmann,W. *et al.* (2004) Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. *J. Exp. Med.*, **200**, 47–59.
21. Bardwell,P.D., Woo,C.J., Wei,K., Li,Z., Martin,A., Sack,S.Z., Parris,T., Edelmann,W. and Scharff,M.D. (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat. Immunol.*, **5**, 224–229.
22. Peron,S., Metin,A., Gardes,P., Alyanakian,M.A., Sheridan,E., Kratz,C.P., Fischer,A. and Durandy,A. (2008) Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. *J. Exp. Med.*, **205**, 2465–2472.
23. Gardes,P., Forveille,M., Alyanakian,M.A., Aucouturier,P., Ilencikova,D., Leroux,D., Rahner,N., Mazerolles,F., Fischer,A., Kracker,S. *et al.* (2012) Human MSH6 deficiency is associated with impaired antibody maturation. *J. Immunol.*, **188**, 2023–2029.
24. Lindahl,T. (1982) DNA repair enzymes. *Annu. Rev. Biochem.*, **51**, 61–87.
25. Hughes,M.J. and Jiricny,J. (1992) The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities. *J. Biol. Chem.*, **267**, 23876–23882.
26. Jiricny,J. (2006) The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.*, **7**, 335–346.
27. Krokan,H.E. and Bjoras,M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.*, **5**, a012583.
28. Neddermann,P. and Jiricny,J. (1994) Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 1642–1646.
29. Nilsen,H., Haushalter,K.A., Robins,P., Barnes,D.E., Verdine,G.L. and Lindahl,T. (2001) Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J.*, **20**, 4278–4286.
30. Hendrich,B., Hardeland,U., Ng,H.H., Jiricny,J. and Bird,A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, **401**, 301–304.
31. Visnes,T., Doseth,B., Pettersen,H.S., Hagen,L., Sousa,M.M., Akbari,M., Otterlei,M., Kavli,B., Slupphaug,G. and Krokan,H.E. (2009) Uracil in DNA and its processing by different DNA glycosylases. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **364**, 563–568.
32. Robertson,A.B., Klungland,A., Rognes,T. and Leiros,I. (2009) DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell. Mol. Life. Sci.*, **66**, 981–993.
33. Frosina,G., Fortini,P., Rossi,O., Carrozzino,F., Raspaglio,G., Cox,L.S., Lane,D.P., Abbondandolo,A. and Dogliotti,E. (1996) Two pathways for base excision repair in mammalian cells. *J. Biol. Chem.*, **271**, 9573–9578.

34. Schanz,S., Castor,D., Fischer,F. and Jiricny,J. (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 5593–5598.
35. Pena-Diaz,J., Bregenhorn,S., Ghodgaonkar,M., Follonier,C., Artola-Boran,M., Castor,D., Lopes,M., Sartori,A.A. and Jiricny,J. (2012) Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol. Cell*, **47**, 669–680.
36. Langerak,P., Nygren,A.O., Krijger,P.H., van den Berk,P.C. and Jacobs,H. (2007) A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J. Exp. Med.*, **204**, 1989–1998.
37. Delbos,F., De Smet,A., Faili,A., Aoufouchi,S., Weill,J.C. and Reynaud,C.A. (2005) Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.*, **201**, 1191–1196.
38. Mayorov,V.I., Rogozin,I.B., Adkison,L.R. and Gearhart,P.J. (2005) DNA polymerase eta contributes to strand bias of mutations of A versus T in immunoglobulin genes. *J. Immunol.*, **174**, 7781–7786.
39. Stavnezer,J. and Schrader,C.E. (2006) Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends Genet.*, **22**, 23–28.
40. Schrader,C.E., Guikema,J.E., Linehan,E.K., Selsing,E. and Stavnezer,J. (2007) Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J. Immunol.*, **179**, 6064–6071.
41. Raschle,M., Dufner,P., Marra,G. and Jiricny,J. (2002) Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J. Biol. Chem.*, **277**, 21810–21820.
42. Fischer,F., Baerenfaller,K. and Jiricny,J. (2007) 5-Fluorouracil is efficiently removed from DNA by the base excision and mismatch repair systems. *Gastroenterology*, **133**, 1858–1868.
43. Grunstein,M. and Hogness,D.S. (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 3961–3965.
44. Pham,P., Bransteitter,R., Petruska,J. and Goodman,M.F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, **424**, 103–107.
45. Shen,H.M. and Storb,U. (2004) Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 12997–13002.
46. Cejka,P., Stojic,L., Mojas,N., Russell,A.M., Heinimann,K., Cannavo,E., di Pietro,M., Marra,G. and Jiricny,J. (2003) Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J.*, **22**, 2245–2254.
47. Iaccarino,I., Marra,G., Dufner,P. and Jiricny,J. (2000) Mutation in the magnesium binding site of hMSH6 disables the hMutSalpha sliding clamp from translocating along DNA. *J. Biol. Chem.*, **275**, 2080–2086.
48. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.
49. Pluciennik,A., Dzantiev,L., Iyer,R.R., Constantin,N., Kadyrov,F.A. and Modrich,P. (2010) PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16066–16071.
50. Casellas,R., Nussenzweig,A., Wuerffel,R., Pelanda,R., Reichlin,A., Suh,H., Qin,X.F., Besmer,E., Kenter,A., Rajewsky,K. *et al.* (1998) Ku80 is required for immunoglobulin isotype switching. *EMBO J.*, **17**, 2404–2411.
51. Manis,J.P., Gu,Y., Lansford,R., Sonoda,E., Ferrini,R., Davidson,L., Rajewsky,K. and Alt,F.W. (1998) Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J. Exp. Med.*, **187**, 2081–2089.
52. Yoshikawa,K., Okazaki,I.M., Eto,T., Kinoshita,K., Muramatsu,M., Nagaoka,H. and Honjo,T. (2002) AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science*, **296**, 2033–2036.
53. Okazaki,I.M., Kinoshita,K., Muramatsu,M., Yoshikawa,K. and Honjo,T. (2002) The AID enzyme induces class switch recombination in fibroblasts. *Nature*, **416**, 340–345.
54. Bowers,P.M., Horlick,R.A., Neben,T.Y., Toobian,R.M., Tomlinson,G.L., Dalton,J.L., Jones,H.A., Chen,A., Altobell,L. 3rd, Zhang,X. *et al.* (2011) Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20455–20460.
55. Greisman,H.A., Lu,Z., Tsai,A.G., Greiner,T.C., Yi,H.S. and Lieber,M.R. (2012) IgH partner breakpoint sequences provide evidence that AID initiates t(11;14) and t(8;14) chromosomal breaks in mantle cell and Burkitt lymphomas. *Blood*, **120**, 2864–2867.
56. Schrader,C.E., Linehan,E.K., Mochegova,S.N., Woodland,R.T. and Stavnezer,J. (2005) Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J. Exp. Med.*, **202**, 561–568.
57. Schrader,C.E., Guikema,J.E., Wu,X. and Stavnezer,J. (2009) The roles of APE1, APE2, DNA polymerase beta and mismatch repair in creating S region DNA breaks during antibody class switch. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **364**, 645–652.
58. Rada,C., Di Noia,J.M. and Neuberger,M.S. (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*, **16**, 163–171.
59. Shen,H.M., Tanaka,A., Bozek,G., Nicolae,D. and Storb,U. (2006) Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *J. Immunol.*, **177**, 5386–5392.
60. Basu,U., Meng,F.L., Keim,C., Grinstein,V., Pefanis,E., Eccleston,J., Zhang,T., Myers,D., Wasserman,C.R., Wesemann,D.R. *et al.* (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, **144**, 353–363.
61. Zarrin,A.A., Del Vecchio,C., Tseng,E., Gleason,M., Zarin,P., Tian,M. and Alt,F.W. (2007) Antibody class switching mediated by yeast endonuclease-generated DNA breaks. *Science*, **315**, 377–381.
62. Flores-Rozas,H. and Kolodner,R.D. (2000) Links between replication, recombination and genome instability in eukaryotes. *Trends Biochem. Sci.*, **25**, 196–200.
63. Paul,A. and Paul,S. (2014) The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Front. Biosci.*, **19**, 605–618.
64. Robbiani,D.F., Bunting,S., Feldhahn,N., Bothmer,A., Camps,J., Deroubaix,S., McBride,K.M., Klein,I.A., Stone,G., Eisenreich,T.R. *et al.* (2009) AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell*, **36**, 631–641.
65. Bardelli,A., Cahill,D.P., Lederer,G., Speicher,M.R., Kinzler,K.W., Vogelstein,B. and Lengauer,C. (2001) Carcinogen-specific induction of genetic instability. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5770–5775.

Fig. S1 Bregenhorn, *et al.*

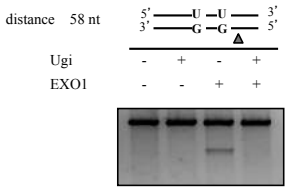
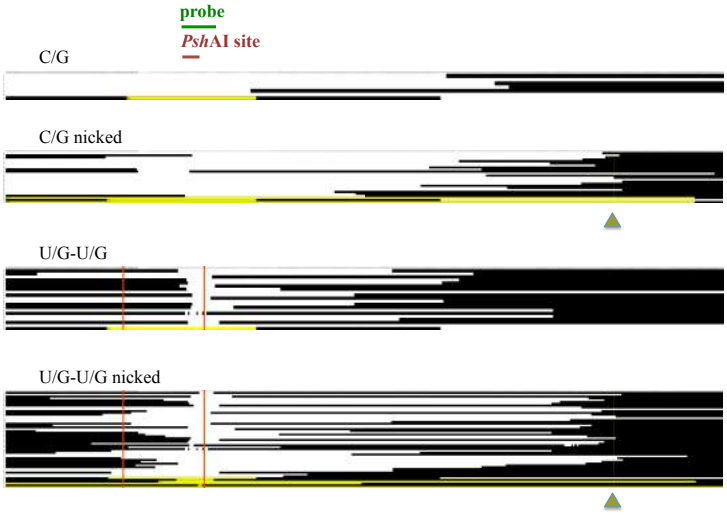


Figure S1. DSB induction in the "58" substrate is EXO1-dependent
The U/G-U/G substrate "58" nicked 5' from the mispaired Gs was incubated with extracts of HEK293 cells pretreated with siRNA targeting EXO1 mRNA and supplemented where indicated with Ugi and/or recombinant purified EXO1. As shown, no DSBs are formed in the absence of EXO1, even though the nick in the bottom strand and the nearest UNG/APE1-cleaved abasic site in the upper strand are only 58 base pairs apart. The figure is a UV-shadowing image of a 0.8% agarose gel stained with GelRed.

Bregenhorn, *et al.*

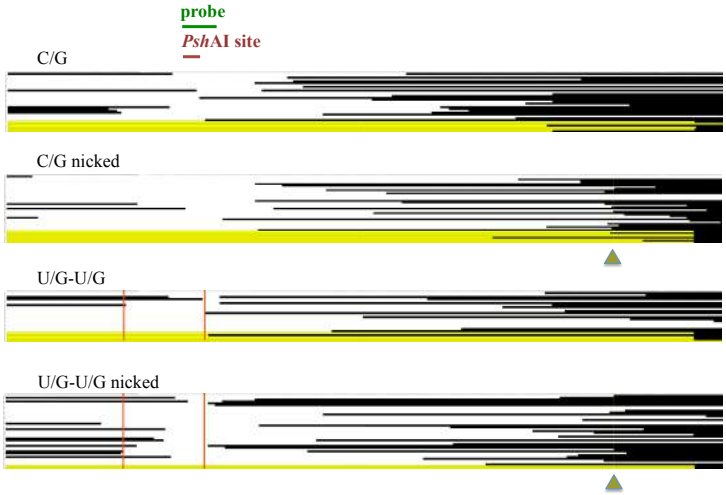
A

HEK 293 wt



B

HEK 293 siMsh2&6



Bregenhorn, *et al.*

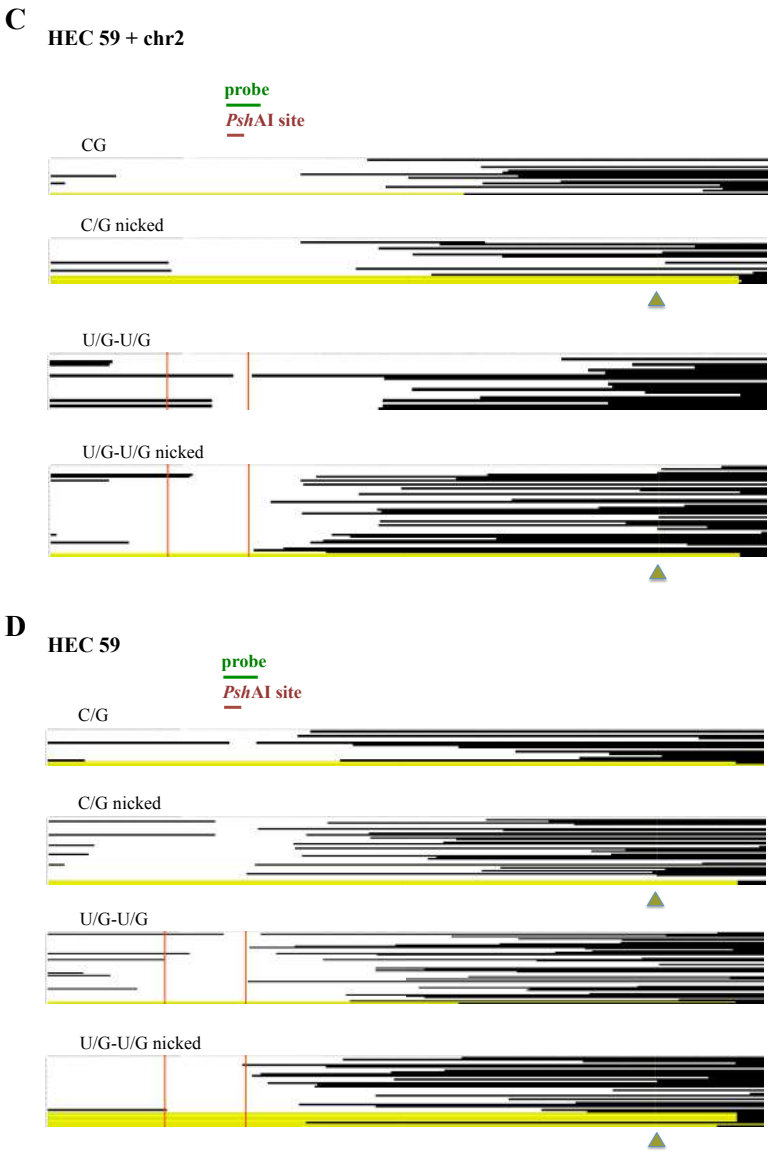


Figure S2. Graphic representation of *in vivo* deletions detected in the U/G-U/G "304" substrate nicked 5' from the mispaired Gs.
A. Deletion in plasmids recovered from multiple transfections of HEK 293 cells with the listed substrates. **B.** As in A, but the substrates were transfected into HEK 293 cells depleted of MutSα. **C.** As in A, but the substrates were transfected into HEC 59 + chr2 cells (MMR-proficient). **D.** As in A, but the substrates were transfected into HEC 59 cells. Vertical red lines indicate the positions of the uracils and the triangle shows the position of the nick.

RESULTS II

A methylation-responsive E-box binding protein controls estrogen-dependent transcription of the chicken vitellogenin gene

Lia Kallenberger[1,3], Rachel Erb[1,3], Josef Jiricny[1,2,3]

[1]Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[2]Department of Biology, Swiss Federal Institute of Technology (ETH) Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[3]Present address: Department of Biology, Institute of Biochemistry, Swiss Federal Institute of Technology (ETH) Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abstract

The regulation of vitellogenin (*VTG*) transcription has been extensively studied. The hormone-inducible expression of the gene was shown to be accompanied by extensive chromatin remodeling and strand-specific DNA demethylation. The pathways involved in the selective removal of 5-methylcytosine (5mC) remained elusive, as there was no plausible mechanism described for active DNA demethylation, despite the common consensus of its existence. With the discovery of the ten-eleven translocation (TET) proteins and their oxidizing activity towards 5mC, a persuasive way to actively demethylate DNA was emerging. We set out to investigate the dynamics of 5mC metabolism in the *VTG* regulatory region in a hepatoma cell line. We saw no demethylation, despite the fact that the gene was efficiently induced upon β-estradiol treatment. We did, however, observe increased levels of 5-hydroxymethylcytosine (5hmC), a product of the TET enzymes. Estrogen receptor α binding to its response element in the *VTG* enhancer was not altered by the presence of 5mC or 5hmC, as assessed by electrophoretic mobility shifts. In nuclear extracts of different origins, however, we saw the binding of a methylation-sensitive factor. *In vitro* methylation of a reporter plasmid under the control of the *VTG* regulatory region inhibited transcription, confirming an important role of methylation dynamics in the control of *VTG* expression. We have evidence that another CpG, residing in an E-box in the VTG promoter, is responsible for the observed methylation-sensitivity. The factor(s) involved in the methylation-dependent transcriptional regulation of *VTG* remain to be characterized and will provide valuable insights into the role of methylation and its removal in the control of gene expression.

Introduction

Methylation of the 5-position of cytosines in the CpG sequence context is a key epigenetic modification of DNA in higher eukaryotes. In vertebrates, around 90% of CpGs are methylated, primarily in the bodies of genes (Eckhardt et al., 2006; Rakyan et al., 2004). Unmethylated CpGs are found in gene regulatory regions, promoters and enhancers. Methylation patterns are established during early embryogenesis by the DNA methyltransferases DNMT1, DNMT3a and DNMT3b. DNMT3a/b are believed to be the major *de novo* methyltransferases that modify unmethylated DNA, while DNMT1 is believed to be the enzyme that maintains

the methylation patterns following replication or repair (Bacolla et al., 1999; Bestor, 1992; Hsieh, 1999; Okano et al., 1998; Pradhan et al., 1999). The essential nature of DNA methylation is demonstrated by the fact that knock-out mouse models show early lethality (Li et al., 1992; Okano et al., 1999).

While the establishment of methylated cytosines is relatively well understood, little is known about the active erasure of methylation patterns. With the discovery of the ten-eleven translocation (TET) enzymes, a new attractive model of active DNA demethylation emerged (Tahiliani et al., 2009). TET enzymes are capable of oxidizing 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) and the latter two have been shown to be excised by TDG and replaced with unmodified cytosine (He et al., 2011; Ito et al., 2011; Zhang et al., 2012). Nevertheless, there are numerous remaining open questions concerning targeting of a possible demethylation complex to specific loci, the exact mechanism of action and the consequences on gene transcription.

DNA methylation in gene regulatory sequences has been generally associated with transcriptional silencing. Many housekeeping genes have CpG-rich regions, the so-called CpG islands, flanking their transcription start sites, which are generally unmethylated and transcriptionally active (Bird et al., 1985; Cooper et al., 1983). Their methylation is mainly described in imprinting, retroviral silencing and X chromosome inactivation, where it is accompanied by gene silencing (Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997). There are also few examples of CpG island methylation at selected loci in a highly cell type specific manner during development (De Smet et al., 1996; De Smet et al., 1999). Established methylation patterns of CpG islands are thought to be relatively stable in somatic cells after development is completed. Aberrant changes are reported in aging and during tumorigenesis, leading to deregulated expression of the affected genes (Baylin et al., 1986; Gama-Sosa et al., 1983; Goelz et al., 1985; Noreen et al., 2014; Toyota et al., 1999). Research on DNA methylation has been heavily focusing on CpG islands, but evidence pointing towards more flexible and dynamic changes of DNA methylation apart from CpG islands is accumulating. These changes are reported to occur in somatic cells upon exogenous stimuli and are also linked to the transcriptional silencing of genes (Amenya et al., 2016; Kangaspeska et al., 2008; Metivier et al., 2008; Thomassin et al., 2001; Toker et al., 2013).

A relatively well-studied example is the chicken vitellogenin (*VTG*) gene that is expressed in the liver of mature hens, but not roosters. This difference correlates with methylation of the enhancer sequence, which contains the estrogen-response element (ERE). Interestingly, the *VTG* gene in rooster liver could be transcriptionally activated by a single β-estradiol (E_2) injection, followed by demethylation of a Hpa II site at the 5' end region of the gene (Wilks et al., 1982). This was accompanied by the appearance of DNaseI hypersensitive sites in the enhancer and promoter (Burch and Weintraub, 1983). Church & Gilbert sequencing of the genomic DNA showed that the transcription was activated after 6 h and that this event coincided with the demethylation of four CpGs in the non-transcribed strand close to the ER binding site. After 24 h, both strands were demethylated. Passive demethylation through replication was excluded at the early time point (Saluz et al., 1986). The molecular mechanism underlying this

phenomenon has not been elucidated to date. In 1986, the molecular machinery involved in DNA demethylation had not been characterized yet.
We therefore wanted to make use of this inducible, well-studied system to address the enzymology behind the DNA methylation and demethylation in hormone-dependent gene activation and the effects on transcription.

Material and Methods

Cell culture
LMH/2A cells were grown in Williams' E medium (GIBCO) without phenol red, supplemented with 10% FCS, streptomycin/penicillin (100 U/ml) and L-Glutamine (2.4 mM). MCF-7 cells were grown in DMEM (GIBCO) with phenol red, supplemented with 10% FCS, streptomycin/penicillin (100 U/ml) and insulin (Sigma-Aldrich, 10 µg/ml). Cell lines were grown at 37°C in a 6% CO_2 humidified atmosphere. Where indicated, cells were treated with 100 nM β-estradiol (E_2, Sigma-Aldrich) or an equivalent volume of ethanol, or with 8 µM aphidicolin (aph, ready made solution, Sigma-Aldrich) or an equivalent volume of DMSO. E_2 and aphidicolin were added freshly every 12 h, because of instability under cell culture conditions. For the TET2 overexpression experiments, cells were transfected with the mouse TET2 expression vector (FH-Tet2-pEF a kind gift from Anjana Rao (Addgene plasmid # 41710)(Ko et al., 2010)) 24 h before E_2 treatment.

Western Blot (WB)
Cells were collected using trypsin, washed in phosphate-buffered saline (PBS) and lysed in RIPA buffer (50 mM Tris pH 8, 1 mM EDTA, 1% NP-40, 0.5% deoxycholate (DOC), 0.1% SDS, 150 mM NaCl). Lysates were sonicated (18 s, 50 cycles, 70% amplitude; in a Bandelin Sonoplus GM70) and protein concentrations were measured using the Bradford assay. 1 µl of sample was diluted in 800 µl H2O and 200 µl of Bradford solution (BioRad) and absorbance was measured at 595 nm (Varian-Cary 50 Scan spectrophotometer). Protein concentration was calculated using a calibration curve with increasing concentrations of bovine serum albumin (BSA). 30-80 µg of protein were boiled in 1x SDS loading buffer (50 mM Tris, pH 6.8, 10% glycerol, 1.6% SDS, 0.1 M DTT, 0.01% bromophenol blue) and separated on polyacrylamide gels (6-10%) using the Mini Trans-Blot Electrophoretic Transfer Cell (BioRad) in 10% SDS-running buffer at 130 V. Proteins were transferred onto activated PVDF membranes (Amersham Pharmacia Biotech) from 1.5 h to overnight in 25 mM Tris, 192 mM glycine and 10% methanol. Membranes were stained with Ponceau S solution (5% acetic acid, 0.1% Ponceau S, Applichem) for 5 min to check equal loading and successful transfer, and blocked for 1 h in 5% non-fat dry milk in 1x TBS-T (20 mM Tris pH 7.4, 150 mM NaCl, 0.1% Tween-20). Membranes were incubated with primary antibody overnight at 4°C in 5% milk, washed 3x with TBS-T and incubated for 1 h at RT in a secondary antibody (horseradish peroxidase-conjugated sheep anti-mouse or donkey anti-rabbit IgG, GE Healthcare). Membranes were washed 3x in TBS-T, incubated in WesternBright Chemiluminescent Detection Reagent (Advansta) for 2 min and analyzed with a Fusion Solo (Vilber Lourmat). Antibodies used were mouse Anti-Vtg (Abcam,

ab36794), rabbit Anti-TFIIH p89 (Santa Cruz, sc-293), mouse Anti-ERα (ThermoFisher, MA5-13065), mouse Anti-TET2 (Abiocode, M1086-1), mouse Anti-Flag M2 (Sigma, F3165), mouse Anti-HA Tag (GenScript, A01244).

Real-time quantitative PCR (RT-qPCR)
Cells were harvested using trypsin and RNA was extracted according to manufacturer's instructions (RNeasy, Qiagen). 2 µg of the extracted RNA was reverse transcribed using the high capacity cDNA reverse transcription kit (Applied Biosystems) according to manufacturer's instructions. 125 ng of the cDNA was used for the PCR using the platinum SYBR green qPCR superMix-UDG kit (Invitrogen) according to manufacturer's instructions, except that the reaction was scaled down from 50 µl to 20 µl. The standard cycling program for ABI instruments was used (50°C for 2 min, 95°C for 2 min, 40x (95°C for 15 s, 60°C for 30 s) and melting curve analysis at 65°C for 15s, heating to 97°C with continuous acquisition per 5°C, 40°C for 30 s). The RT-qPCR was run on a LightCycler 480 (Roche). GapDH primers were used as an internal control. Technical triplicates were made for every sample and primer. The primers used are listed in Table S1 (SI).

Chromatin immunoprecipitation
75 Mio LMH/2A cells were harvested after 24 h treatment with 100 nM E_2 or EtOH. Cells were washed twice with PBS and fixed for 10 min in 1% formaldehyde at RT. Fixation was quenched with 125 mM glycine at RT for 10 min. Cells were collected by centrifugation at 300 x g for 5 min at 4°C. Cells were washed twice with PBS and the pellet was resuspended in cell lysis buffer (10 mM Tris, pH 8, 1 mM EDTA, 0.5% IGEPAL, protease inhibitors) and incubated on ice for 10 min. Lysate was centrifuged at 2000 x g at 4°C for 5 min. The pellet was resuspended in nuclear lysis buffer (10 mM Tris, pH 8, 1 mM EDTA, 0.5 M NaCl, 1% Triton X-100, 0.5% sodium dexycholate, 0.5% lauroylsarcosine, protease inhibitors) and incubated on ice for 10 min with repeated vortexing. Lysate was centrifuged at 3000 x g for 5 min at 4°C. The pellet was resuspended in PBS, split for different immunoprecipitations and centrifuged at 3000 x g for 10 min at 4°C. Pellets were resuspended in 300 µl lysis buffer (10 mM Tris, pH 8, 1 mM EDTA, 150 mM NaCl, 0.1% sodium deoxycholate, 0.1% SDS, protease inhibitors) and sonicated at maximum power for 10 min with 30 s intervals in a Bioruptor (Diagnode). Samples were centrifuged at 16'000 x g for 10 min at 4°C and the supernatant was collected. Input samples (15%) were frozen and the rest was made up to 1 ml with IP buffer (16.7 mM Tris, pH 8, 1.2 mM EDTA, 300 mM NaCl, 1.1% Triton X-100, protease inhibitors). 5 µg of the respective antibodies (Anti-ERα; ThermoFisher MA5-13065 or Anti-Flag; Sigma F3165) were added to the lysate and incubated on a rotating wheel at 4°C overnight.
Beads (G protein Sepharose, GE Healthcare) were washed twice in IP buffer and centrifuged for 2 min at 2700 x g before being blocked with 1 mg/ml BSA in IP buffer for 1 h. After one additional wash, 50 µl of beads were added to 1 ml lysate with antibody. The mixture was incubated for 3 h on a rotating wheel at 4°C. Mixtures were centrifuged for 2 min at 2000 x g and the supernatant was removed. The beads were washed successively in 1 ml of four different wash buffers for 5 min on the rotating wheel (wash buffer 1: 20 mM Tris, pH 8, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 300 mM NaCl; wash buffer 2: like wash buffer

1, but with 500 mM NaCl; wash buffer 3: 10 mM Tris, pH 8, 1 mM EDTA, 1% sodium deoxycholate, 500 mM LiCl, 1% NP-40; wash buffer 4: 10 mM Tris, pH 8, 1 mM EDTA). One additional wash with wash buffer 4 was performed, during which the beads were split for DNA and proteins. For proteins: the beads were taken up in 2x SDS loading dye (100 mM Tris, pH 6.8, 20% glycerol, 3.2% SDS, 0.2 mM DTT, 0.04% bromophenol blue), the mixture was vortexed vigorously and incubated for 30 min at 100°C (denaturing and de-cross-linking). The beads were spun down and the supernatant was loaded on an SDS-PAGE together with inputs. Gels were processed further according to general WB protocol. For DNA: beads and inputs were taken up in 100 µl crosslink reversal buffer (20 mM Tris, pH 8, 0.5 mM EDTA, 0.1 M $NaHCO_3$, 1% SDS, RNase at 10 µg/ml (Roche)) and incubated at 65°C overnight. The next day, the DNA was purified using the PCR purification kit from Qiagen and eluted in 10 µl of H_2O. RT-PCR was performed on 3.5 µl of DNA using the platinum SYBR green qPCR superMix-UDG kit (Invitrogen) according to manufacturer's instructions, except that the reaction was scaled down to 20 µl and 1 mM of $MgCl_2$ were added. The following protocol was used for amplifications: 50°C for 2 min, 95°C for 2 min, 50x (95°C for 15 s, 57°C for 30 s) and melting curve analysis at 65°C for 15 s, heating to 97°C with continuous acquisition per 5°C, 40°C for 30 s). The qPCR was run on a LightCycler 480 from Roche with the following primers: fwd 5'-TAAGCAAAGCTTTGGCTGCATTTTGTATTTGGTTCTGTAGGAAATG-3', rev 5'-TAAGCAGCTAGCCTGCAGGCCAGAGGCTAATCCAGATAAACC-3'.

Labeling of oligos
Fill-in reaction (3'-labeling): 2 pmoles of ds 5'-overhanged DNA were incubated for 15 min at RT with 0.2 µl of the respective [α-^{32}P]-dNTP (2 µCi, Hartman Analytic), 1 mM of each of the other three dNTPs and 5 U Klenow fragment (3'—>5' exo-, NEB) in the corresponding buffer. The enzyme was heat-inactivated for 20 min at 75°C and slowly cooled down to re-anneal the strands. Free dNTPs were removed on a Sephadex G-25 column (GE Healthcare).

Kinase reaction (5'-labeling): 2 pmoles of ssDNA were phosphorylated with 0.7 µl of [γ-^{32}P]-ATP (7 µCi, Hartman Analytic) using T4 polynucleotide kinase (10 U, NEB). The reaction was supplemented with 10 mM DTT and incubated for 30 min at 37°C. The reaction was heat-inactivated for 5 min at 95°C and free nucleotides were removed on a Sephadex G-25 column (GE Healthcare). The oligo was annealed to a 1.5x excess of unlabeled complementary strand by heating for 5 min to 80°C and slowly cooled down in annealing buffer (50 mM Hepes, pH 7.5, 100 mM NaCl).

Nuclear extracts
All nuclear extracts were prepared according to Dignam et al. (Dignam et al., 1983), with or without pretreatment of the cells with E_2.

Electrophoretic Mobility Shift Assay (EMSA)
All EMSA reactions were carried out in binding buffer (10 mM Tris-HCl, pH 7.6, 50 mM NaCl, 1 mM DTT, 1 mg/ml BSA, 5% glycerol (Hyder et al., 1999)) in a volume of 5 µl using the unspecific competitor poly(dI-dC) (Sigma-Aldrich). 20 ng of poly(dI-dC) was used for 100 ng of recombinant ERα (ThermoFisher, RP-

310) and 1 µg for 10 µg of nuclear extracts. For all EMSAs, 10 fmol of [α-^{32}P]-dNTP labeled, or [γ-^{32}P]-ATP phosphorylated oligos were used. Where indicated, an excess of unlabeled competitor DNA or 100 nM of E_2 were added. For the supershift assays, 200 ng of antibody (Anti-ERα; ThermoFisher MA5-13065) were added together with the proteins. Proteins, and competitor or antibodies where indicated, were incubated for 20 min on ice in binding buffer. Labeled DNA was added and the mixture was left at RT for another 20 min before loading on a 5% polyacrylamide gel (Acryl/ Bis 29:1, Amresco) eluted with 1x TAE (40 mM Tris, 20 mM acetate, 1 mM EDTA). The gels were run at 200 V for 1 h in 1x TAE and dried in a gel dryer for 1 h. The gels were exposed to phosphor screens and the autoradiographs were developed in a typhoon FLA 9500 (GE Healthcare Life Sciences). The oligonucleotides used are listed in Table S2 (SI).

SDS-PAGE of cross-linked binding reactions

EMSAs were performed as described and cross-linked for 5 min (~720 mJoules) in a UV Stratalinker 1800 (Stratagene). Samples were taken up in 2x SDS loading buffer and boiled for 5 min before being loaded on a 7.5% polyacrylamide gel and run for 1-2 h in 10% SDS-running buffer at 130 V. Gels were dried for 1 h, exposed to phophor screens and developed in a typhoon FLA 9500 (GE Healthcare Life Sciences).

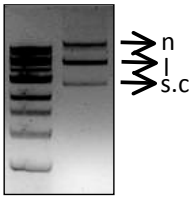
CpGL cloning and Luciferase assay

The enhancer/promoter region of the chicken vitellogenin gene spanning nucleotides -638 to -1 was cloned into a CpG-free firefly luciferase plasmid (pCpGL-basic, InvivoGen) using *Hind*III and *Nco*I. The enhancer promoter region was amplified from genomic DNA of LMH/2A cells using the forward primer 5'-TAAGCAAAGCTTTGGCTGCATTTTGTATTTGGTTCTGTAGGAAATG-3' and the reverse primer 5'-ATTCGTCCATGGCGAAGGTGAATAGGGCTCTGCC-3'. The plasmid was further cut with *Hind*III and *Hpa*II to excise the intervening sequence and the annealed upper 5'-AGCTTAAAAATATTCCTGGTCAGCGTGAC-3' and lower 5'-CGGTCACGCTGACCAGGAATATTTTTA-3' ds-oligo was ligated into the plasmid to translocate the *Hind*III site closer to the ERE for further ligations of the ERE. On the other side of the ERE a *Kpn*I site was created by mutagenesis using the following primers: new *Kpn*I fwd 5'-CCGGAGCTGAAAGAACACATGGTACCCGTGATTTCAATAAATAC-3'; new *Kpn*I rev 5'-GTATTTATTGAAATCACGGGTACCATGTGTTCTTTCAGCTCCGG-3'. Another *Kpn*I site in the plasmid had to be destroyed by mutagenesis: destroy KpnI site fwd 5'-CATGTTCACAACACTGGGGTATCTGATCTGTGGCTTCAGAG-3'; destroy KpnI site rev 5'-CTCTGAAGCCACAGATCAGATACCCCAGTGTTGTGAACATG-3'.

The resulting VTG-CpGL plasmid was co-transfected with a Renilla luciferase-expressing plasmid (pRL-SV40, Promega) into LMH/2A cells using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Cells were re-seeded the next day in technical triplicates both for E_2 and ethanol treatment and treated for 24 h. Cells were subsequently lysed directly in firefly luciferase substrate from Promega (Dual-Glo Luciferase kit) for 30 min, with vigorous shaking. Cell lysates were processed further according to manufacturer's instructions and luminescence was measured on the SpectraMax i3 (Molecular Devices) plate reader. Relative luciferase units (RLU) were calculated as the ratio between firefly and Renilla signals.

Mutagenesis of VTG-CpGL
200 ng of vector DNA were mixed with 50pmoles forward and reverse primers, 250 µM dNTPs, 2% DMSO and 1 U phusion high fidelty polymerse (NEB) in 1x buffer provided by the manufacturer. Extension was performed in 50 µl according to the following protocol: 95 °C, 2 min; 30x (95 °C, 1 min, 55 °C, 1 min, 63 °C, 30 min); 68 °C, 20 min; 15 °C hold. Extension reaction was digested twice with 20 U DpnI for 1 h at 37°C followed by heat inactivation of the enzyme for 20 min at 80°C. The reaction mix was then transformed into electro-competent Pir1 bacteria (R6K gamma ORI, Invitrogen) after being desalted by incubating for 15 min on a 0.025 µm Millipore nitrocellulose filter on H_2O. Bacteria were shaken at 37°C for 40 min and plated onto Zeocin (25 µg/ml, InvivoGen) agar plates. Clones were picked the next day, grown in liquid culture overnight and plasmids were extracted using the NucleoSpin Plasmid Kit (Macherey-Nagel) according to manufacturer's instructions. Plasmids were sequenced at Microsynth to check for successful mutagenesis. The primers for mutagenesis are listed in Table S3 (SI). For sequencing, the following primers were used: VTG enhan seq fwd 5'-GCAATCAATATTGAAAACCAC-3'; CpGL seq luc1 fwd 5'-CGTGAAGAGAATTTATGATTGC-3'; CpGL seq luc2 fwd 5'-CCCTGACACTGCCATCCTG-3'.

Ligation of modified ERE into VTG-CpGL
50 µg of the plasmid DNA were cut with *Hind*III-HF (NEB) in Cut Smart buffer with 100 U of enzyme for 1.5 h at 37°C in 50 µl. Linearization was verified on a 1% agarose gel and enzyme was heat-inactivated for 20 min at 80°C. For the first ligation step, 10 µg of linearized DNA and 20x excess of annealed oligo with different modifications at the CpGs (Upper: 5'-AGCTGAAAAATATTCCTGGTCAGCGTGACCGGAGCTGAAAGAACACATGGTAC-3'; Lower: 5'-CATGTGTTCTTTCAGCTCCGGTCACGCTGACCAGGAATATTTTTC-3') were incubated for 2 h at RT in T4 ligase buffer with 400 U of T4 ligase (NEB) in a total volume of 100 µl. Ligase was inactivated by heating to 65°C for 10 min and efficient ligation was verified on an agarose gel. Reaction volume was increased to 200 µl with 1x Cut Smart buffer, 40 U of *Kpn*I-HF (NEB) were added and incubation was continued for additional 1.5 h at 37°C. 20 U of *Hind*III-HF were then added, the mixture was incubated for an additional hour and subsequently cleaned-up on MinElute columns (5µg per column, Qiagen) and eluted in 20 µl H_2O. 20 reactions of the first ligation were pooled for the second ligation. 12.5 µg of DNA from the first ligation were incubated in 10 ml of 1x T4 ligase buffer with 4000 U of ligase for 1 h at RT. Recircularization was verified on an agarose gel.



n, nicked. l, linear. sc, super-coiled

The resulting mix was concentrated by EtOH precipitation and the supercoiled form of the plasmid was extracted on a CsCl gradient and purified further as described (Baerenfaller et al., 2006).

In vitro methylation
1 µg of plasmid DNA was incubated in 1x NEB buffer 2, 160 µM freshly diluted SAM (NEB) and 4 U of *Sss*I (NEB) for 1 h at 37°C. Reaction was stopped by heating to 65°C for 20 min.

EdU incorporation and Click-iT Reaction
Cells were grown on coverslips and treated with 8 µM aphidicolin or DMSO, 100 nM E_2 or ethanol and 10 µM of 5-ethynyl-2'-deoxyuridine (EdU, Invitrogen) for 24h. Medium was removed, cells were washed once with PBS and 1 ml of 3.7% formaldehyde in PBS was added. Cells were fixed for 15 min at RT before being washed twice with 1 ml of 3% BSA in PBS. 1 ml of 0.5% Triton X-100 in PBS was added and cells were incubated for 20 min at RT. Click-iT reaction master mix was prepared (for 5 slides): 129 µl 1x Click-iT reaction buffer (freshly diluted 1:10 in H_2O), 6 µl $CuSO_4$, 0.36 µl Alexa Fluor azide, 15 µl reaction buffer additive (freshly diluted 1:10 in H_2O). Permeabilization buffer was removed and cells were washed twice with 1 ml 3% BSA in PBS. 30 µl of Click-iT reaction mix were pipetted onto Parafilm and the coverslips were put cells-down onto the mix and incubated for 30 min at RT, protected from light. Coverslips were washed once with 1 ml of 3% BSA in PBS and then washed well with 1x PBS to remove BSA. After one final wash with H_2O, cells were fixed with mounting media containing 4',6-diamidino-2-phenylindole (DAPI, VectaShield). Slides were analyzed on an Olympus IX81 fluorescence microscope.

DNA extraction, oxidation of hmC and bisulfite conversion
Genomic DNA was extracted from cells using the Wizard Genomic DNA Purification Kit (Promega). DNA was eluted with water and digested with *Eco*RV overnight. EtOH precipitation was performed and DNA was additionally cleaned up on Micro Bio-Spin 6 chromatography columns in SSC (BioRad) as purity was essential for oxidation of 5hmC. Samples were split and one half was subjected to oxidation. Selective oxidation of 5hmC to 5C was achieved using potassium perruthenate ($KRuO_4$, Sigma-Aldrich) as described (Booth et al., 2012; Booth et al., 2013). In short, 0.5-2 µg of DNA were incubated in 50 mM NaOH in 24 µl for 30 min at 37°C after vigorous vortexing to denature DNA. 1 µl of 15 mM $KRuO_4$ solution in 50 mM NaOH was added and oxidation was incubated on ice for 1 h with vortexing every 5 min. Reaction was cleaned up on polyacrylamide columns (89849, ThermoScientific) and processed further with the rest of the sample for bisulfite conversion. Bisulfite conversion was achieved using the EZ DNA Methylation-Gold Kit from Zymo Research according to manufacturer's instructions. The cycling protocol was adapted because of the slightly less efficient conversion of fC as compared to unmodified cytosine (95°C for 5 min; 2x (60°C for 25 min, 95°C for 5 min, 60°C for 85 min, 95°C for 5 min, 60°C for 175 min, 95°C for 5 min); 20°C hold).

Polymerase chain reaction on bisulfite converted DNA
100 ng of bisulfite-converted DNA was used as template for the PCR using the ZymoTaq DNA polymerase (Zymo Research). The reaction was carried out in a total volume of 40 µl in the reaction buffer provided by the manufacturer supplemented with of 1 mM dNTPs, 1.5 mM $MgCl_2$, 500 mM each primer and 2 U of polymerase. The cycling protocol was as follows: 10 min 95°C; 5x (30 s 94°C; 30 s 52°C; 90 s 72°C); 5x (30 s 94°C; 30 s 52°C; 90 s 72°C); 35x (30 s 94°C; 30 s 55°C; 90 s 72°C); 7 min 72°C. The primers used were: Upper strand forward 5'-TAAGCAAAGCTTTTGTTTATTTAAAAATATTTTTGGTTAG-3'; Upper strand reverse 5'-TAAGCAGCTAGCCCTCAACTTTAATTTATAACCATTC-3'; Lower strand forward 5'-TAAGCAAAGCTTAGTTTTGGTTTATGGTTATTTTTGTA-3'; Lower strand reverse 5'-TAAGCAGCTAGCCAAAATACTTATAAAACCTTTATTCATTTA-3'. The forward primers were only added after the first 5 cycles to reduce the formation of primer dimers. The PCR fragments were purified using a BluePippin (Sage Science) on a 2% agarose gel according to manufacturer's instructions, before being bar-coded for PacBio sequencing.

PacBio sequencing and analysis
Pacific Biosciences developed a new sequencing approach that uses single-molecule real-time technology (SMRT), referred to as PacBio sequencing. The converted RSII sequencers allow time-resolved single-molecule studies based on custom experimental design (in manual operation mode). Any desired probe can be immobilized on the surface of the SMRT cell, which consists of 150'000 zero-mode waveguide (ZMW) nanostructures allowing to work with up to 50'000 single molecules in parallel. The nanostructures allow for discrimination between single molecule binding events and fluorescent background even at high analyte concentration.
SMRT is a fluorescence-based method. For each of the 150'000 ZMW nanostructures, traces in four wavelength channels are stored. Software for data evaluation was created by informaticians at the Functional Genomic Center Zurich (FGCZ). The technology is able to produce long reads *via* a strand-displacement mechanism of the polymerase. The resulting high-quality circular consensus reads faithfully correct the high error-rate of the polymerase, resulting in very accurate sequencing results.

Results

The vitellogenin enhancer is methylated and the gene silenced in LMH/2A cells prior to exposure to β-estradiol and is induced upon hormone treatment, independently of DNA replication.
In a first step, we wanted to test whether the phenomenon described by Saluz et al. (Saluz et al., 1986) was reproducible and whether E_2 would activate the silenced VTG gene also in a system more amenable to manipulation. We decided to use chicken LMH/2A cells, a Leghorn rooster hepatocellular carcinoma cell line stably expressing estrogen receptor alpha (ERα). We could show by bisulphite sequencing that the four CpGs in the VTG enhancer (a-d, Fig. 1 A), described by Saluz et al., are about 90% methylated, except for CpG b, which is 54% methylated in both strands (Fig. 1 B). This is in line with the results of Saluz

et al. from rooster liver, where the four CpGs were also methylated except for CpG b that was described to be 50% hemimethylated. With our approach we cannot say with certainty whether all of CpG b is hemimethylated or half is un - and half is fully-methylated, but the latter is much more likely, pointing towards a loss of hemimethylation in the cell line.

Using RT-qPCR, we could show that the *VTG* gene is not transcribed in LMH/2A cells prior to exposure to estrogen, but is efficiently induced already 6 h after addition of 100 nM E_2 and expression increases further after 24 h (Fig. 1 C). The increase in transcription continues up to 72 h as shown by RT-PCR and Western Blot (Supp. Fig. 1 A), at least partly due to a positive feedback resulting in increased transcription of estrogen receptor α (Supp. Fig. 1 B).

Activation of *VTG* transcription occurs in a replication-independent manner, as inhibition of DNA replication by addition of the B-family polymerase inhibitor aphidicolin to the experiment does not impede transcription, but rather enhances it (Fig. 1 D). A possible explanation for this observation is that any interference of replication with the transcription process is inhibited by the addition of aphidicolin. The successful inhibition of DNA synthesis was confirmed with a 5-ethynyl-2'-deoxyuridine (EdU) incorporation assay (Supp. Fig. 1 C).

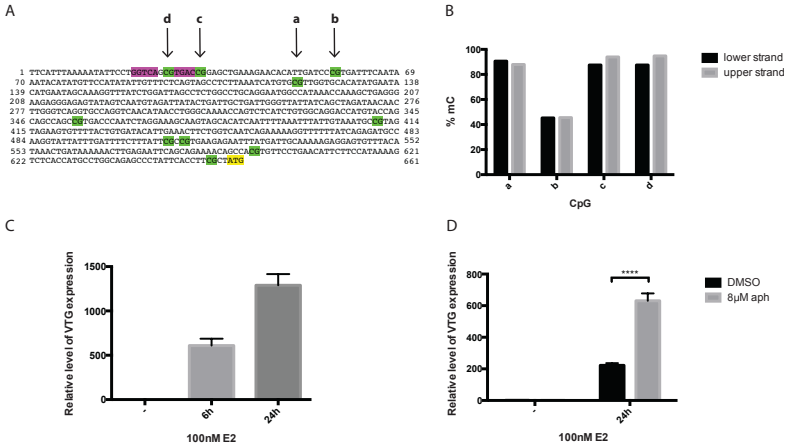


Fig 1: VTG mRNA expression is induced upon β–estradiol (E2) treatment in LMH/2A cells, independently of DNA synthesis. A Sequence of the enhancer/promoter region of the chicken *VTG* gene. The ERE binding site (violet), the CpGs (green) and the translation start site (yellow) are highlighted. The four CpGs (**a-d**) analyzed in Saluz *et al.* are indicated . **B** Bisulfite sequencing results of the four CpGs in the *VTG* enhancer region for the lower and the upper DNA strand. **C** *VTG* mRNA levels were measured with qPCR after 6 and 24 hours of 100 nM E2 treatment and with additional treatment with 8 µM aphidicolin or DMSO for 24h (**D**). Data are represented as mean ± SD. Significance was assessed using Sidak's multiple comparisons test.

Estrogen receptor alpha binds to the ERE of VTG upon treatment with β-estradiol.

Next we checked whether the estrogen receptor physically interacts with the ERE in the VTG enhancer upon E_2 treatment. To this end we performed a chromatin immunoprecipitation experiment using an ERα antibody or a flag antibody as a control. We could retrieve ERα in the chromatin fraction only with the ERα antibody and with E_2 treatment (Fig. 2, upper panel). Moreover, we

could confirm its binding to the VTG ERE using RT-qPCR on the recovered DNA (Fig. 2, lower panel). Only very little signal was apparent without E_2 treatment, confirming the widely accepted model that estrogen receptors bind DNA *in vivo* only in response to hormone treatment (Carroll et al., 2005; Klinge, 2001).

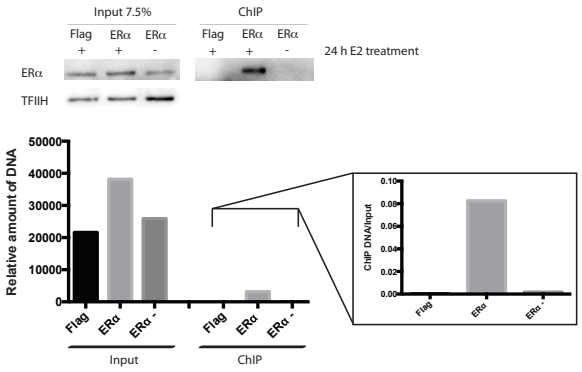


Fig 2: ERα directly interacts with the *VTG* enhancer upon E2 treatment. Upper panel: Western Blot of chromatin extracts immunoprecipitated with Flag- or ERα antibodies (- and + β–estradiol treatment). Lower panel: RT-qPCR of immunoprecipitated DNA. Right panel: Ratio of signal of immunoprecipitated DNA and input DNA.

The four CpGs in the enhancer region of VTG are not demethylated in LMH/2A cells, but there is evidence for elevated levels of 5hmC upon E_2 treatment.

After showing that VTG expression is induced upon hormone treatment and ERα binds to the enhancer, we set out to investigate if demethylation or TET-mediated oxidation of the involved CpGs took place. Therefore, we performed PCR on the bisulfite-converted DNA to distinguish C, 5fC and 5caC from 5mC and 5hmC. To separate mC from hmC signals, bisulfite conversion was preceded by an oxidation step using $KRuO_4$, which selectively oxidizes 5hmC to 5fC, resulting in its conversion to uracil following bisulfite treatment. The PCR products were then sequenced using the PacBio approach, which is able to provide information on single molecules. 5hmC levels were assessed as the difference in the amount of converted cytosines between bisulfite only - and oxidizing-bisulfite converted DNA.

Using this approach, we could show that the promoter/enhancer was not demethylated during the course of the induction, but that a small percentage of 5mCs were oxidized to 5hmCs, indicating an involvement of one or several of the TET enzymes (Fig. 3 A and Supp. Fig. 2). TET 1 and 2 were shown to be present in LMH2A cells as assessed by qPCR and WB, as well as all cytosine modifications as assessed by Dot Blot (data not shown). We next checked the effect of TET2 overexpression in LMH/2A cells on the induction of VTG expression and found a significant increase in the level of induction upon mouse TET2 overexpression (Fig. 3 B). Successful overexpression was confirmed by Western Blot using Anti-TET2, -Flag and -HA antibodies (Fig. 3 C).

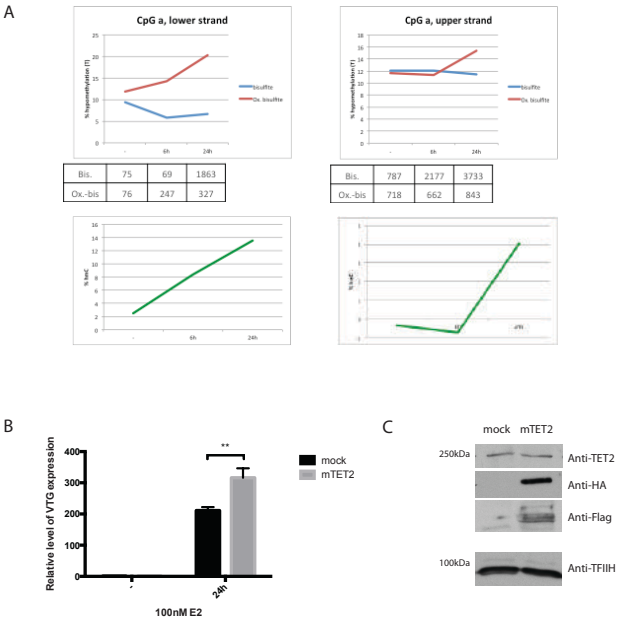


Fig 3: CpGs a-d are not demethylated, but E2 treatment brings about conversion of 5mC to 5hmC. **A** Representative bisulfite and oxidated-bisulfite sequencing result of CpG **a**. The numbers below the upper panel indicate the total amount of analyzed reads per condition. The percentage of 5hmC (lower panel) was calculated as the difference of converted Cs between the oxidated-bisulfite converted DNA and the only-bisulfite converted DNA. **B** Induction of *VTG* mRNA with mock treatment or mTET2 overexpression measured by qPCR (left panel). Data are represented as mean ± SD. **C** Western Blot showing overexpression of Flag-HA-tagged mTET2.

Binding of ERα to the ERE is insensitive to different cytosine modifications and hormone treatment.

Since it was published that transcriptional activation goes hand in hand with demethylation of the ERE and we saw elevated levels of hmC, we next wanted to assess whether binding of the ERα was dependent on the methylation status of the ERE. EMSA assays using synthetic oligonucleotides containing the VTG ERE wt or ΔG, missing one nucleotide in the 3-nt spacer between the palindromic repeats (Fig. 4 A) and recombinant ERα (ThermoFisher, RP-310) showed that the recombinant protein was able to bind the ERE sequence with similar affinity in presence and absence of E_2, on unmethylated (Fig. 4 B & C) and methylated oligos (Fig. 4 B, lower panel). Moreover, the mutant ERE ΔG bound ERα with significantly reduced affinity (Fig. 4 B, last lanes). The binding of ERα to the ERE was competed with a 100-fold excess of specific competitor, but not with a scrambled sequence (Fig. 4 C, lanes 4 & 5). In addition, there was a change in conformation upon addition of E_2, resulting in altered mobility, confirming the functionality of the receptor (Fig. 4 C & Supp. Fig. 3 A). Supershift experiments using an ERα antibody further confirmed the specificity of the shift (Fig. 4 D)

The binding of ERα could be shown to be insensitive to methylation of the two CpGs within the ERE at different receptor concentrations (Fig. 4 E). Hemimethylation (Supp. Fig. 3 B & C), as well as hydroxy- (Supp. Fig. 3 B & D) and hemihydroxymethylation (Supp. Fig. 3 B & E), of the ERE also failed to affect ERα binding. Also formylcytosine and carboxycytosine in the ERE had no influence on receptor binding (Supp. Fig. 3 B). To confirm the specificity of the binding, another mutant missing all three nucleotides of the spacer, resulting in direct palindromic repeats, was included and competition assays were performed on the 45 bp oligos and additionally on shorter, 25 bp oligos (Supp. Fig. 3 F & G). The methylated and the hydroxymethylated oligos successfully competed with the unmethylated oligos, whereas the mutants did not.

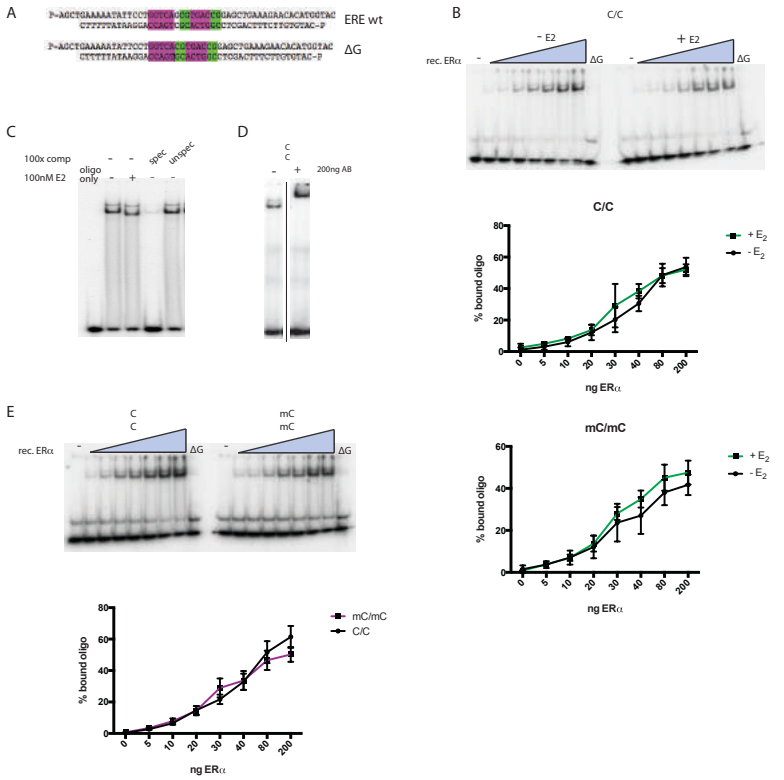


Fig 4: Binding of recombinant ERα to the ERE is methylation- and hormone-insensitive. A Oligos used for radioactive EMSAs, showing the ERE (violet) and the CpGs (green). **B** Upper panel: Representative EMSA with increasing concentrations of recombinant ERα (0, 5, 10, 20, 30, 40, 80 and 200 ng) on unmethylated ERE wt oligo and ΔG oligo in the last lanes with 40 ng of ERα, - (left panel) /+ (right panel) E2. Quantification of percentage bound oligo on unmethylated (middle panel) or methylated (lower panel) oligo of three independent experiments. Data are represented as mean ± SD. **C** EMSA with 60 ng recombinant ERα and unmethylated ERE wt oligo, with or without E2 or competitor. **D** EMSAs showing a supershift with ERα antibody with unmethylated oligo. **E** Upper panel: Representative EMSA with increasing concentrations of recombinant ERα (0, 5, 10, 20, 30, 40, 80 and 200 ng) and unmethylated (left panel) or methylated (right panel) ERE wt oligo and ΔG oligo in the last lanes with 40 ng of ERα. Lower panel: Quantification of percentage bound oligo of three independent experiments. Data are represented as mean ± SD.

In vitro methylation of a luciferase reporter plasmid under the control of the VTG enhancer/promoter inhibits transcription.

In the study performed by Saluz et al., transcriptional activation with hormone treatment was accompanied by changes in methylation of the enhancer. Even though we saw no evidence of a direct impact of ERE methylation on receptor binding, we were interested to check the impact of different methylation states on transcription. We generated a reporter plasmid in a CpG-free backbone, expressing firefly luciferase under the control of the *VTG* enhancer/promoter (Fig. 5 A). We than replaced the ERE with synthetic oligonucleotides either unmethylated or methylated at the two CpGs or *in vitro* methylated the whole construct, resulting in methylation of all CpGs in the enhancer/promoter region. The substrates were transfected into LMH/2A cells and cells were treated with either 100 nM E_2 or EtOH for 24h.

The analysis of unmethylated and *in vitro* methylated wt and ΔG plasmid showed impaired expression when the ERE is mutated, but the plasmid was still inducible by E_2 (Fig. 5 B).

Both plasmids show very weak transcriptional activity and hormone-inducibility when *in vitro* methylated (Fig. 5 B). Next we wanted to check whether methylation of the CpGs in the ERE was responsible for the inhibition of transcription. Therefore, we included a plasmid *in vitro* methylated with *Hpa*II, affecting only CpG c (according to Saluz et al.) in the ERE, and saw no effect of this methylation (Fig. 5 C). Using a ligation-mediated approach, we then exchanged the ERE region containing two CpGs (c and d) with an oligo containing mC or hmC at both CpGs. The luciferase assay showed no difference in transcriptional activity between these different modifications (Fig. 5 D). This is in line with the EMSA results showing that binding of ERα is unaffected by different modifications, but it opens the question why the *in vitro* methylated plasmids have impaired transcriptional activity even though ERα is able to bind to its response element. The next steps therefore were to investigate whether the binding behavior of ERα is different in nuclear extracts compared to the recombinant protein and/or if methylation of another CpG in the enhancer/ promoter is responsible for the inhibition of transcription.

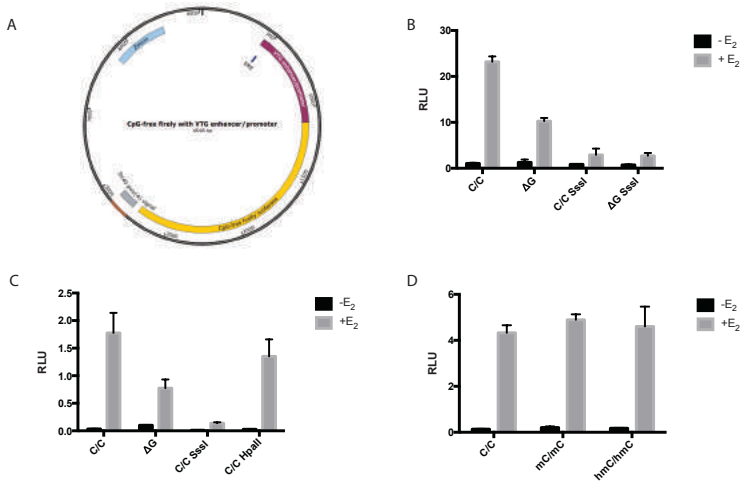


Fig 5: Transcription of a luciferase reporter is inhibited upon methylation of the *VTG* enhancer/promoter region. **A** Scheme of the CpG-free luciferase reporter under the control of the *VTG* enhancer/promoter. **B** Luciferase expression of unmethylated and *in vitro* methylated (*Sss*I) CpG-free plasmid with wt ERE (C/C) and a single nucleotide deletion in the spacer of the ERE (ΔG). **C** As **B**, additionally with a plasmid methylated with *Hpa*II, only present in one CpG in the ERE. **D** Luciferase expression from the plasmid with different cytosine modifications only in the two CpGs in the ERE. Relative luciferase units (RLU) are defined as ratio between firefly and Renilla signal. The graphs show the mean ± SD of three independent experiments.

Another protein binds the ERE wt and ΔG in nuclear extracts

In order to investigate whether the binding characteristics of ERα in nuclear extracts are the same as of the recombinant ERα, we prepared nuclear extracts (NEs) according to Dignam et al. (Dignam et al., 1983) of LMH/2A cells as well as MCF-7 and Hela cells and carried out EMSAs with the ERE oligos. We could detect a shifted double band with the unmethylated wt oligo in LMH/2A NE, which was even stronger with the ΔG oligo. The band with higher electrophoretic mobility was methylation-sensitive. The two bands migrated slightly faster than the complex formed by recombinant ERα and there was no change in mobility when E_2 was added to the mixture (Fig. 6 A). The same behavior was seen in MCF-7 NE (Fig. 6 B). Hemi-methylated oligos retained the band with higher electrophoretic mobility, whereas hemi -and fully-hydroxymethylated oligos lost it similar to fully methylated oligos (Supp. Fig. 4 A). Supershift assays using an ERα antibody had no effect on the shifts (Supp. Fig. 4 B). These observations prompted us to conclude that the protein binding the wt and mutant ERE was not ERα, but another, yet unknown protein (referred to as factor X). To narrow down the binding sites for additional proteins, shorter oligos were used for all further experiments (Supp. Fig. 3 G & Fig. 6 C). The shifts on the shorter oligos showed slightly altered mobility, but the methylation-sensitivity was still detectable. Direct comparison of binding to ERE of recombinant ERα and NE clearly showed the different behavior, most strikingly on the ΔG oligo that bound ERα with strongly reduced affinity and factor X with higher affinity as compared to wt ERE (Fig. 6 C). Retinoic acid (RA) and retinoic X (RX) receptors are known

to bind to similar sequences as the ERs (Mader et al., 1993). To check this possibility we used a competitor oligo shown to bind RAR and RXR and saw no competition in contrast to competition with ΔG (Supp. Fig. 4 C).
Since we were not able to see an ERα shift in the NEs, we wanted to assess the effect of proteins present in the NE on binding of the recombinant ERα. Therefore we titrated recombinant ERα into the binding reactions with NE. Recombinant ERα bound wt ERE unmethylated and methylated also together with NEs, but with reduced affinity, most likely due to competition with other proteins such as factor X (Fig. 6 D). ΔG that was bound very strongly by factor X did not bind any ERα titrated to the reaction. Quantification of the binding of ERα versus factor X of three independent experiments revealed that the methylated oligo bound ERα with higher affinity compared to factor X than the unmethylated oligo (Fig. 6 D lower panel) also pointing towards a competition of binding of ERα versus factor X rather than a synergy.
We then set out to characterize factor X further. Therefore binding reactions with NEs were performed on oligos containing one BrdU on each strand (Fig. 6 E) and cross-linked by UV. Reactions were then resolved on SDS-PAGE gels to assess the exact size of the bound proteins. We were able to detect the expected band of the ERα-oligo complex of about 80 kDa in LMH/2A NE on un-, hemi- and fully-methylated oligos, which displayed similar mobility like the recombinant protein cross-linked to the oligo (Fig. 6 E). In addition, there was another band visible at around 50 kDa on all oligos and a band at around 170 kDa if the upper (labeled) strand was methylated. Subtracting the weight of the single stranded oligo (roughly 15 kDa) this corresponded to proteins of about 30-40 kDa and 150-160 kDa. When we titrated the recombinant receptor into the reaction with NE we saw a weak, but reproducible competition with the smaller protein (Fig. 6 F). The binding of recombinant ERα on the unmethylated oligo could be outcompeted with specific, but not with unspecific competitor (Fig. 6 F, last three lanes). There was no effect upon addition of E_2 on the binding affinities or the ratios of binding of the different proteins (Fig. 6 G). Also the ERα band was unchanged, confirming that the shift with estradiol in the EMSA was a conformational change that is not visible in the denaturing conditions of the SDS gel. The pattern of the bound proteins looks similar in MCF-7 NE, except for an additional low molecular weight band around 55 kDa (Fig. 6 H).

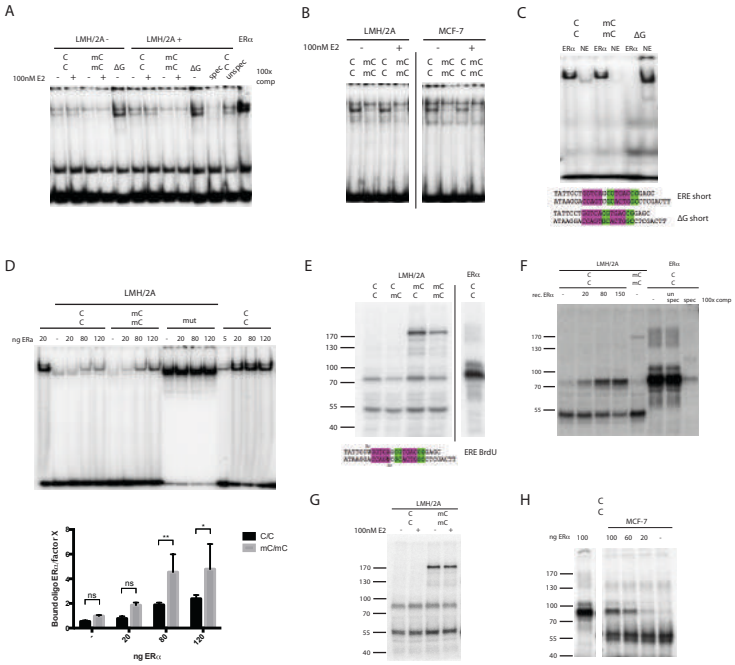


Fig 6: A methylation-sensitive shift on ERE wt and ΔG that is not mediated by ERα. **A** EMSA with LMH/2A nuclear extracts (NEs) pretreated for 24 h with EtOH (LMH/2A-) or 100 nM E2 (LMH/2A+) and recombinant ERα (last lane). Where indicated, reaction mixtures were supplemented with EtOH (-), 100 nM E2 (+) or 100 x excess of specific (spec) or unspecific (unspec) competitor. **B** EMSA comparing binding of LMH/2A and MCF-7 NEs on unmethylated and methylated ERE oligo, -/+ E2. **C** EMSA comparing recombinant protein with LMH/2A NEs on shorter oligos (ERE short) unmethylated (C/C) and methylated (mC/mC) and a shorter ΔG oligo (ΔG short). **D** Upper panel: EMSA with LMH/2A NE with titration of indicated amounts of recombinant ERα on different oligos. Lower panel: Quantification of ratio of bound oligo by ERα and factor X of three independent experiments. Data are represented as mean ± SD. Significance was assessed using the Holm-Sidak test. For **A** and **B** oligos from Fig 4 A were used. For **C** and **D** oligos depicted in **C** were used. **E-H** UV-crosslinked binding reaction using 60 µg of LMH/2A or MCF-7 (**H**) NE and recombinant ERα and short oligos containing BrdU (**E**, lower panel). Where indicated, 100 nM E2, recombinant ERα or competitor was added to the reaction mix. Crosslinked complexes were resolved on an 8% polyacrylamide gel and visualized on phosphor screens.

CpG7 shows a methylation-sensitive shift in LMH/2A, MCF-7 and HeLa extracts.
Since we observed a strong inhibitory effect of methylation in the enhancer/promoter region of VTG that was shown to be independent of methylation of the ERE, we set out to investigate the other CpGs in this region to find the CpG conferring methylation-dependent inhibition. The eight CpGs (following the two in the ERE) were numbered from one to eight, CpG b from Saluz et al. being CpG1, CpG a being CpG2 (Fig. 1 A). CpG5 and 6 were analyzed together as they lie very close to each other. We performed EMSAs with NEs on these CpGs and were surprised to see the same shift with different intensities on the ERE, CpG1, 2 and 7 (Fig. 7 A). Another molecular mobility shift was seen on CpG3 (Supp. Fig. 4 D) and no shifts on CpG4, 5/6 and 8 in LMH/2A as well as MCF-7 NEs (Fig. 7 A and Supp. Fig. 4 E). Interestingly, the shift of seemingly the same protein on the ERE, CpG1, 2 and 7 showed distinct methylation sensitivities

on the different sequences. The methylation sensitivity on the ERE has been seen before and also on CpG1 there was an influence of methylation, but not on CpG2. The strongest effect of methylation however, could be seen on CpG7. The same behavior was visible in MCF-7 (Supp. Fig. 4 E) and in HeLa extracts (Supp. Fig. 4 F). SDS gels with cross-linked binding reactions showed a band with similar mobility on CpG7 as on the ERE and an additional band migrating slightly slower in LMH/2A (Fig. 7 B) and in HeLa NE (Fig. 7 C).

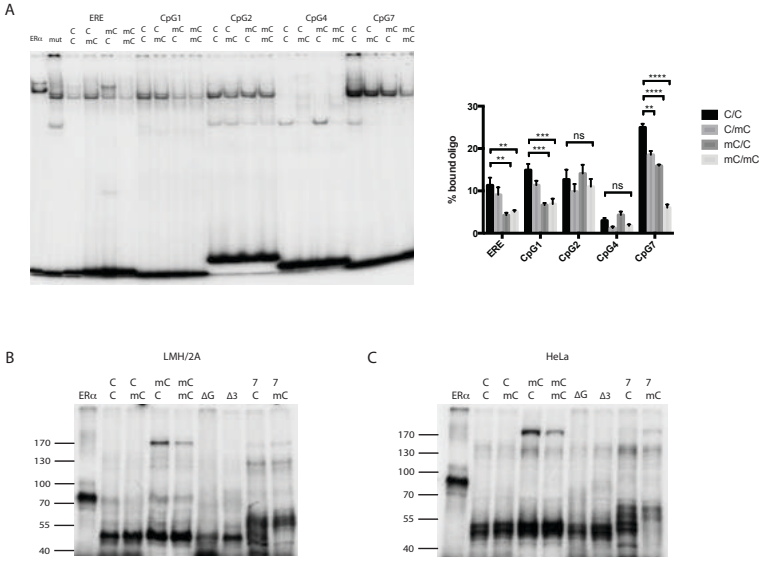


Fig 7: CpG 7 shows strong methylation-sensitivity of a complex migrating at the same height as on the ERE.
A Left panel: EMSA with LMH/2A nuclear extracts on four different CpG-containing regions bearing different cytosine combinations. Right panel: Quantification of percentage bound oligo of three independent experiments. Data are represented as mean ± SD. Significance was assessed using the Tukey test for multiple comparisons.
B & C UV-crosslinked binding reaction using 60 µg of LMH/2A (**B**) or HeLa (**C**) NE and recombinant ERα (first lanes) and short oligos containing BrdU (**Fig. 6 E**, lower panel). Oligo Δ3 misses the 3nt of the spacer region in the ERE. Crosslinked complexes were resolved on an 8% polyacrylamide gel and visualized on a phosphor screen.

CpG7 plays an important role in the methylation-sensitivity of the expression of luciferase under control of the VTG enhancer/promoter.
We next set out to characterize the behavior of the different CpGs further using the luciferase assay. Since it is very elaborate to introduce methylated oligos for all CpGs using ligation, we decided to mutate the CpGs one by one from CpG to TpG, therefore inhibiting their methylation by *Sss*I (Fig. 8 A). First we checked whether the C to T transition itself influenced the expression of luciferase, impeding the investigation of methylation-sensitivity. We found that mutation at the ERE (mut ERE) and CpG7 (C594T) were the only ones negatively affecting luciferase expression absolutely (Fig. 8 B & C & Supp. Fig. 5 A & B) as well as fold

induction upon E_2 treatment (Supp. Fig. 5 C). They had to be kept unchanged and were as a consequence always methylated when the plasmids were treated with *Sss*I. To check whether the mutation in CpG7 directly affects factor X binding, we performed an EMSA on the mutant oligo and saw reduced binding as compared to the wt unmethylated oligo bearing CpG7, but still stronger binding than with wt methylated CpG7 (Fig. 8 D). This translated directly to luciferase expression, where the mutant CpG7 showed intermediate expression between wt unmethylated and wt methylated, hinting towards an important role of factor X binding in transcriptional activation. The expression in all the other mutants was still inhibited when methylated with *Sss*I, showing that these individual sites, when unmethylated, had no influence on transcription (Fig. 8 E & Supp. Fig. 5 D).

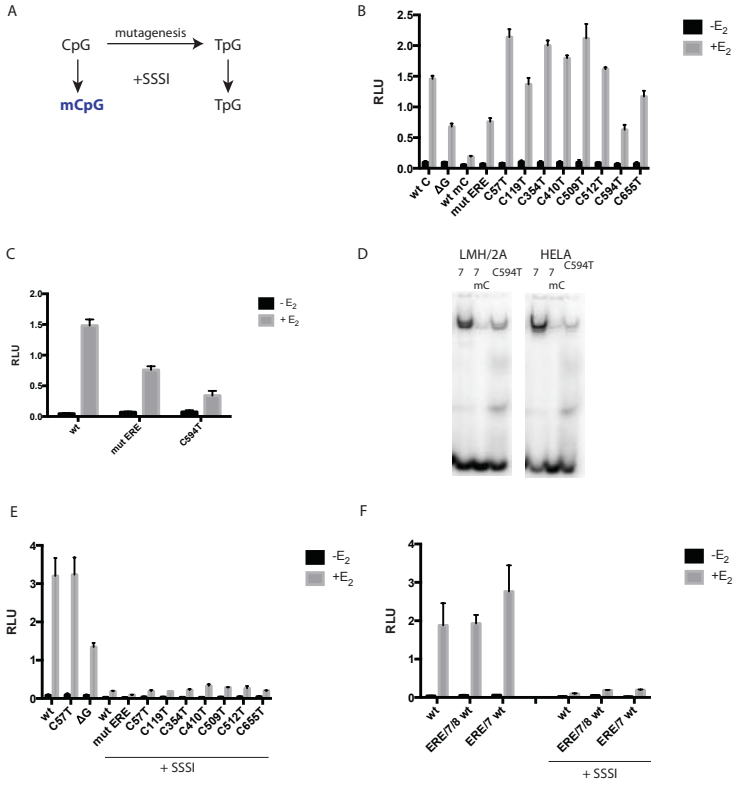


Fig 8: CpG7 is responsible for the methylation-sensitivity of the VTG enhancer/promoter. A Scheme of experimental setup. **B** Luciferase assay of different mutants. One representative experiment is shown. Data are represented as mean of technical triplicates ± SD. **C** Luciferase assay of wt unmethylated plasmid (wt), C to T mutations in the ERE (mut ERE) and mutation of CpG7 (C594T). Data are represented as mean of three independent experiments ± SD. **D** EMSA showing binding of factor X to unmethylated (7), methylated (7 mC) and mutant CpG7 (C594T) in LMH/2A and HeLa nuclear extracts. **E** Same as **B** but mutants are methylated with SSSI. **F** Luciferase assay of wt CpGL, of the CpGL with all CpGs mutated except for the ERE, CpG7 and CpG8 (ERE/7/8 wt) and the CpGL with all mutated except for the ERE and CpG7(ERE/7 wt), unmethylated and methylated (+SSSI). Data are represented as mean of three independent experiments ± SD.

Since CpG7 seems to play an important role in the control of VTG transcription, but we could not analyze the mutant, we had to use a different approach. We set out to mutate all CpGs together, except for CpG c and d in the ERE and CpG7. We tested the combined mutants successively to exclude that interactions between different sites had any additional influence on transcription, which was not the case. All combinations tested were still active when unmethylated and inhibited when methylated (Supp. Fig. 5 E). Analysis of the final plasmid (ERE/7 wt) unmethylated and methylated confirmed that methylation of CpG7 and possibly the combination with methylated ERE (even though methylation of ERE alone had no effect, Fig. 5 D), was conferring the inhibition (Fig. 8 F). This prompted us to conclude that proteins binding to CpG7 were key to the regulation of expression of VTG and that their binding was methylation-sensitive. Figure 9 shows a summary of the results obtained with the luciferase experiments, highlighting again the importance of CpG7 (arrows).

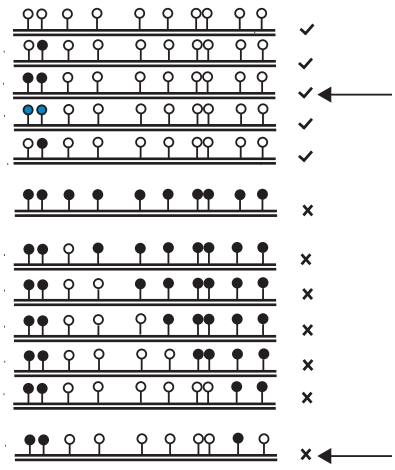


Fig 9: CpG7 is responsible for the methylation-sensitivity of the *VTG* enhancer/promoter. Summary scheme of luciferase assays with the different CpGLs. White circles represent unmethylated, black circles methylated and blue circles hydroxymethylated CpGs. Ticks indicate induction of expression, crosses indicate inhibited expression. The first five lines show plasmids prepared with ligation of short oligos bearing the ERE with different cytosine modifications. The others were *in vitro* methylated using SssI, the unmethylated CpGs were mutated.

Discussion

We could show that vitellogenin expression is induced upon β-estradiol treatment in LMH/2A cells, but that this is happening independently of detectable changes in DNA methylation. The four CpGs reported earlier to become demethylated upon hormone treatment (Saluz et al., 1986) remained resistant to bisulfite conversion. We did, however, see accumulation of the TET oxidation product 5hmC, already 6 h after E_2 treatment and increasing further with longer exposure to E_2. In our system, we cannot distinguish between low expression of *VTG* in all cells or high expression in a subset of cells that underwent demethylation, possibly escaping our analysis, or expression resulting from oxidation of 5mC to 5hmC in some cells. The question why we are unable to detect demethylation remains unresolved, especially because proteins

with an assigned role in demethylation, namely the TETs and thymine DNA glycosylase (TDG), and that are present in chicken embryonic and adult liver, are also present in the cell line. The involvement of yet unknown factors in the demethylation event at the *VTG* enhancer is, however, conceivable.

The finding that 5hmC is enriched at the CpGs upon hormone treatment and the increase in *VTG* expression when TET2 is overexpressed, points towards a functional role of the oxidized base in transcription. The levels of 5hmC are, however, low and there is the possibility that TET2 influences gene expression in a way other than oxidation of 5mC. Another intriguing point is that no oxidation beyond 5hmC was detected. 5fC or 5caC would have led to the bisulfite-dependent conversion, which was not detectable. The question arises whether this is a situation where the TET-mediated oxidation cascade specifically stops after a single step. Another possible explanation would be that excision of 5fC and 5caC and re-methylation of the replaced cytosine are extremely fast and efficient processes. Such cyclical DNA methylation and demethylation events have been observed in the pS2 promoter upon hormone treatment (Kangaspeska et al., 2008; Metivier et al., 2008).

Even though we did not detect demethylation of the enhancer, the observation that we can shut down expression with *in vitro* methylation of the gene regulatory region points towards an important role of CpG methylation in the regulation of *VTG*. It indicates that methylation in this system is more than a mere basal state, resulting as a consequence of gene inactivity before encounter with β-estradiol.

The fact that we do not see differential binding of the estrogen receptor α when the ERE is methylated or hydroxymethylated puzzled us at first, considering the reported importance of methylation at this locus and our observations with *in vitro* methylation. It makes, however, sense that the ERα is able to bind all forms of cytosine modification, as it is the initiating event upon hormone treatment. All chromatin-remodeling processes follow E_2 treatment. Since several nuclear receptors (NRs) were shown to interact with chromatin remodelers and proteins involved in demethylation, such as TDG, stimuli- and site-specific recruitment of a demethylation complex by NRs would constitute an effective system.

We found that the inducibility of the reporter under the control of the *VTG* regulatory region was dependent on E_2 treatment and on methylation, also in the ΔG ERE that did not bind ERα efficiently. Binding of ERα, assessed by EMSAs, however, was independent of both, E_2 treatment and methylation status. It has been shown previously that ERα binds the ERE independently of E_2 *in vitro* (Murdoch et al., 1990). The inducibility *in vivo* seems to be regulated at the level of nuclear import after dissociation from the heat shock protein.

The fact that the reporter with the ΔG ERE was still inducible by E_2, albeit roughly half as efficiently, seemed puzzling. For all we know, the inducibility upon E_2 treatment had to be mediated by the ERα. The VTG regulatory region contains another, imperfect ERE, which could mediate the effect on expression. Moreover, there was a very weak residual binding of ERα to the mutated ERE that possibly also contributes.

The discrepancy between methylation responsiveness of the reporter and the *in vitro* binding of ERα to the ERE can have two reasons. The first could be binding of other proteins to the ERE *in vivo* that confer methylation-sensitivity. Binding of two non-histone proteins (NHPs) to the ERE (NHP1) and its proximity (NHP2)

within the *VTG* enhancer has been described (Feavers et al., 1987). We indeed detected a methylation-sensitive mobility shift (factor X) on the ERE in nuclear extracts (NEs) that was not cell type-specific. We doubt, however, that this protein is similar to the NHPs published, due to different molecular weights as assessed by denaturing SDS gels. Additional evidence arguing against the NHPs are the strongest contact sites of NHP1 that were found to be the AGCG, lying in the spacer region between the palindromic half sites (Hughes et al., 1989). We see a stronger binding of factor X on the ΔG substrate that misses the middle G of this sequence. Another observation made, concerning factor X binding to the ERE, was that this was the only shift we saw in NEs. We do not see any ERα binding, despite the presence of ERα in the NEs, arguing for a competition for the ERE binding site.

The second explanation for the different methylation sensitivities in the reporter and the EMSA could lie in the regulatory sequence downstream of the ERE. There are six additional CpGs in the control region of *VTG*, besides the four CpGs published. The inhibitory effect of methylation was most likely mediated by one or several of these CpGs. We were able to pin down the inhibitory effect of the *in vitro* methylation to CpG number 7, lying in an E-box motif. Methylation sensitivity of helix-loop-helix (HLH) E-box binding proteins has been described previously (Perini et al., 2005; Prendergast et al., 1991; Prendergast and Ziff, 1991) and Prendergast *el al*. additionally reported that different combinations of HLH factors show different binding characteristics. It will be interesting to investigate whether we see a similar phenomenon in the *VTG* regulatory region. The fact that we see the same shift on oligonucleotides covering the different CpGs, but with different affinity and methylation-sensitivity, strongly suggests a complex of different combinations of proteins. This versatile system would endow the cell with great flexibility to its environment using a limited number of transcription factors.

Amenya, H.Z., C. Tohyama, and S. Ohsako. 2016. Dioxin induces Ahr-dependent robust DNA demethylation of the Cyp1a1 promoter via Tdg in the mouse liver. *Scientific reports* 6:34989.

Bacolla, A., S. Pradhan, R.J. Roberts, and R.D. Wells. 1999. Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna. *The Journal of biological chemistry* 274:33011-33019.

Baerenfaller, K., F. Fischer, and J. Jiricny. 2006. Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro. *Methods in enzymology* 408:285-303.

Baylin, S.B., J.W. Hoppener, A. de Bustros, P.H. Steenbergh, C.J. Lips, and B.D. Nelkin. 1986. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. *Cancer research* 46:2917-2922.

Bestor, T.H. 1992. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *Embo J* 11:2611-2617.

Bird, A., M. Taggart, M. Frommer, O.J. Miller, and D. Macleod. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40:91-99.

Booth, M.J., M.R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, and S. Balasubramanian. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336:934-937.

Booth, M.J., T.W. Ost, D. Beraldi, N.M. Bell, M.R. Branco, W. Reik, and S. Balasubramanian. 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature protocols* 8:1841-1851.

Burch, J.B., and H. Weintraub. 1983. Temporal order of chromatin structural changes associated with activation of the major chicken vitellogenin gene. *Cell* 33:65-76.

Carroll, J.S., X.S. Liu, A.S. Brodsky, W. Li, C.A. Meyer, A.J. Szary, J. Eeckhoute, W. Shao, E.V. Hestermann, T.R. Geistlinger, E.A. Fox, P.A. Silver, and M. Brown. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122:33-43.

Cooper, D.N., M.H. Taggart, and A.P. Bird. 1983. Unmethylated domains in vertebrate DNA. *Nucleic Acids Res* 11:647-658.

De Smet, C., O. De Backer, I. Faraoni, C. Lurquin, F. Brasseur, and T. Boon. 1996. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl Acad Sci U S A* 93:7149-7153.

De Smet, C., C. Lurquin, B. Lethe, V. Martelange, and T. Boon. 1999. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Molecular and cellular biology* 19:7327-7335.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11:1475-1489.

Eckhardt, F., J. Lewin, R. Cortese, V.K. Rakyan, J. Attwood, M. Burger, J. Burton, T.V. Cox, R. Davies, T.A. Down, C. Haefliger, R. Horton, K. Howe, D.K. Jackson, J. Kunde, C. Koenig, J. Liddle, D. Niblett, T. Otto, R. Pettett, S. Seemann, C. Thompson, T. West, J. Rogers, A. Olek, K. Berlin, and S. Beck. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature genetics* 38:1378-1385.

Feavers, I.M., J. Jiricny, B. Moncharmont, H.P. Saluz, and J.P. Jost. 1987. Interaction of two nonhistone proteins with the estradiol response element of the avian vitellogenin gene modulates the binding of estradiol-receptor complex. *Proc Natl Acad Sci U S A* 84:7453-7457.

Gama-Sosa, M.A., V.A. Slagel, R.W. Trewyn, R. Oxenhandler, K.C. Kuo, C.W. Gehrke, and M. Ehrlich. 1983. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res* 11:6883-6894.

Goelz, S.E., B. Vogelstein, S.R. Hamilton, and A.P. Feinberg. 1985. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* 228:187-190.

He, Y.F., B.Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.X. Song, K. Zhang, C. He, and G.L. Xu. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303-1307.

Hsieh, C.L. 1999. In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Molecular and cellular biology* 19:8211-8218.

Hughes, M.J., H.M. Liang, J. Jiricny, and J.P. Jost. 1989. Purification and characterization of a protein from HeLa cells that binds with high affinity to the estrogen response element, GGTCAGCGTGACC. *Biochemistry* 28:9137-9142.

Hyder, S.M., C. Chiappetta, and G.M. Stancel. 1999. Interaction of human estrogen receptors alpha and beta with the same naturally occurring estrogen response elements. *Biochem Pharmacol* 57:597-601.

Ito, S., L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, and Y. Zhang. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300-1303.

Kangaspeska, S., B. Stride, R. Metivier, M. Polycarpou-Schwarz, D. Ibberson, R.P. Carmouche, V. Benes, F. Gannon, and G. Reid. 2008. Transient cyclical methylation of promoter DNA. *Nature* 452:112-115.

Klinge, C.M. 2001. Estrogen receptor interaction with estrogen response elements. *Nucleic acids research* 29:2905-2919.

Ko, M., Y. Huang, A.M. Jankowska, U.J. Pape, M. Tahiliani, H.S. Bandukwala, J. An, E.D. Lamperti, K.P. Koh, R. Ganetzky, X.S. Liu, L. Aravind, S. Agarwal, J.P. Maciejewski, and A. Rao. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468:839-843.

Li, E., T.H. Bestor, and R. Jaenisch. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915-926.

Liu, W.M., R.J. Maraia, C.M. Rubin, and C.W. Schmid. 1994. Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res* 22:1087-1095.

Mader, S., P. Leroy, J.Y. Chen, and P. Chambon. 1993. Multiple parameters control the selectivity of nuclear receptors for their response elements. Selectivity and promiscuity in response element recognition by retinoic acid receptors and retinoid X receptors. *The Journal of biological chemistry* 268:591-600.

Metivier, R., R. Gallais, C. Tiffoche, C. Le Peron, R.Z. Jurkowska, R.P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, and G. Salbert. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452:45-50.

Murdoch, F.E., D.A. Meier, J.D. Furlow, K.A. Grunwald, and J. Gorski. 1990. Estrogen receptor binding to a DNA response element in vitro is not dependent upon estradiol. *Biochemistry* 29:8377-8385.

Noreen, F., M. Roosli, P. Gaj, J. Pietrzak, S. Weis, P. Urfer, J. Regula, P. Schar, and K. Truninger. 2014. Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *Journal of the National Cancer Institute* 106:

Okano, M., D.W. Bell, D.A. Haber, and E. Li. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247-257.

Okano, M., S. Xie, and E. Li. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics* 19:219-220.

Perini, G., D. Diolaiti, A. Porro, and G. Della Valle. 2005. In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation.

Proceedings of the National Academy of Sciences of the United States of America 102:12117-12122.

Pradhan, S., A. Bacolla, R.D. Wells, and R.J. Roberts. 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *The Journal of biological chemistry* 274:33002-33010.

Prendergast, G.C., D. Lawe, and E.B. Ziff. 1991. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* 65:395-407.

Prendergast, G.C., and E.B. Ziff. 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* 251:186-189.

Rakyan, V.K., T. Hildmann, K.L. Novik, J. Lewin, J. Tost, A.V. Cox, T.D. Andrews, K.L. Howe, T. Otto, A. Olek, J. Fischer, I.G. Gut, K. Berlin, and S. Beck. 2004. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS biology* 2:e405.

Saluz, H.P., J. Jiricny, and J.P. Jost. 1986. Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis. *Proc Natl Acad Sci U S A* 83:7167-7171.

Tahiliani, M., K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, and A. Rao. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930-935.

Thomassin, H., M. Flavin, M.L. Espinas, and T. Grange. 2001. Glucocorticoid-induced DNA demethylation and gene memory during development. *Embo J* 20:1974-1983.

Toker, A., D. Engelbert, G. Garg, J.K. Polansky, S. Floess, T. Miyao, U. Baron, S. Duber, R. Geffers, P. Giehr, S. Schallenberg, K. Kretschmer, S. Olek, J. Walter, S. Weiss, S. Hori, A. Hamann, and J. Huehn. 2013. Active demethylation of the Foxp3 locus leads to the generation of stable regulatory T cells within the thymus. *Journal of immunology* 190:3180-3188.

Toyota, M., N. Ahuja, M. Ohe-Toyota, J.G. Herman, S.B. Baylin, and J.P. Issa. 1999. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* 96:8681-8686.

Walsh, C.P., J.R. Chaillet, and T.H. Bestor. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics* 20:116-117.

Wilks, A.F., P.J. Cozens, I.W. Mattaj, and J.P. Jost. 1982. Estrogen induces a demethylation at the 5' end region of the chicken vitellogenin gene. *Proc Natl Acad Sci U S A* 79:4252-4255.

Woodcock, D.M., C.B. Lawler, M.E. Linsenmeyer, J.P. Doherty, and W.D. Warren. 1997. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *The Journal of biological chemistry* 272:7810-7816.

Zhang, L., X. Lu, J. Lu, H. Liang, Q. Dai, G.L. Xu, C. Luo, H. Jiang, and C. He. 2012. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol* 8:328-330.

Kallenberger *et al.*, SI

Supplementary Information

RT-qPCR primers	
GapDH fwd	GGCCATCAATGATCCCTTCATCG
GapDH rev	GACAGTGCCCTTGAAGTGTCCG
VTG fwd	CATACTGGCATTAGTGCTCACCC
VTG rev	GGTAACTCCTTCTGCTATTGAATCC
ER1 fwd	CCAGCAGGTACCCTACTACCTTG
ER1 rev	GCTGTGGCGCCTATTATCAGAGC

Table S 1: RT-qPCR primers.

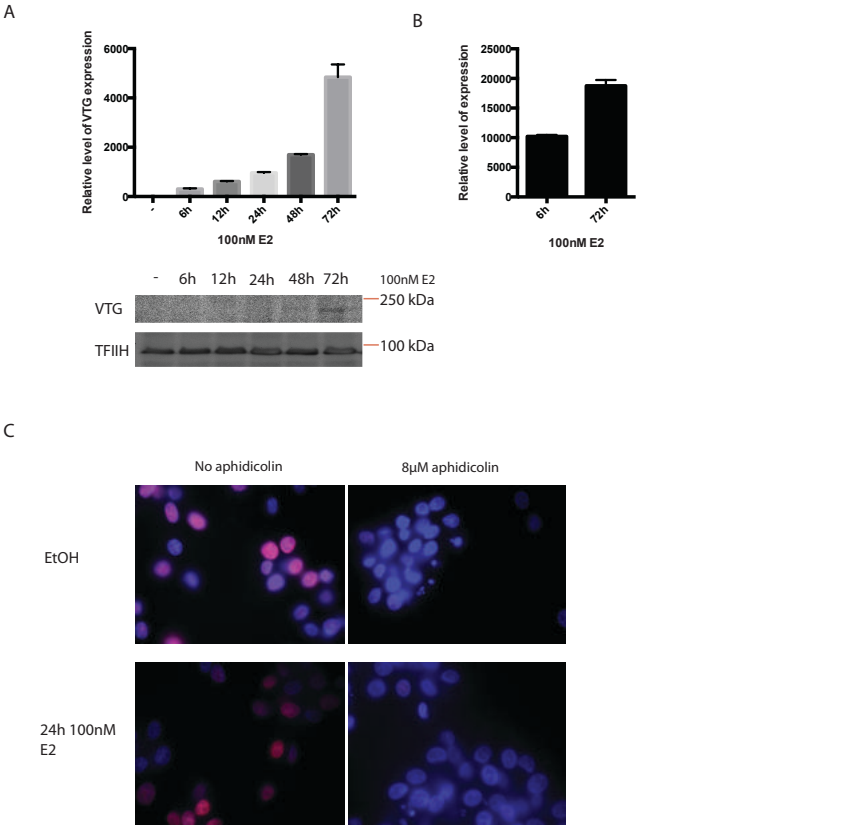
EMSA and pull-down primers	
EMSA comp scr	CTCAAGCTTCCCAACGTCGACAGACGTCTGG
EMSA ERE long up	AGCTGAAAAATATTCCTGGTCAGCGTGACCGGAGCTGAAAGAACAC ATGGTAC
EMSA ERE long low	CATGTGTTCTTTCAGCTCCGGTCACGCTGACCAGGAATATTTTTC
EMSA ERE long ΔG up	AGCTGAAAAATATTCCTGGTCACGTGACCGGAGCTGAAAGAACACA TGGTAC
EMSA ERE long ΔG low	CATGTGTTCTTTCAGCTCCGGTCACGTGACCAGGAATATTTTTC
EMSA ERE long Δ3 up	AGCTGAAAAATATTCCTGGTCATGACCGGAGCTGAAAGAACACATG GTAC
EMSA ERE long Δ3 low	CATGTGTTCTTTCAGCTCCGGTCATGACCAGGAATATTTTTC
EMSA ERE short up	TATTCCTGGTCAGCGTGACCGGAGC
EMSA ERE short low	TTCAGCTCCGGTCACGCTGACCAGGAATA
EMSA ERE short ΔG up	TATTCCTGGTCACGTGACCGGAGC
EMSA ERE short ΔG low	TTCAGCTCCGGTCACGTGACCAGGAATA
EMSA ERE short Δ3 up	TATTCCTGGTCATGACCGGAGC
EMSA ERE short Δ3 low	TTCAGCTCCGGTCATGACCAGGAATA
EMSA CpG1 up	GAACACATTGATCCCGTGATTTCAAT
EMSA CpG1 low	ATTGAAATCACGGGATCAATGT
EMSA CpG2 up	CTTAAATCATGTGCGTTGGTGCACATATGA
EMSA CpG2 low	GTATTCATATGTGCACCAACGCACATGATTTAAG
EMSA CpG3 up	GTACCAGCAGCCAGCCGTGACCCAATCTAG
EMSA CpG3 low	CTAGATTGGGTCACGGCTGGCTGCTG
EMSA CpG4 up	ATTGTAAATGCCGTAGTAGAAGTGTTT
EMSA CpG4 low	GTAAAACACTTCTACTACGGCATTTACAAT
EMSA CpG5_6 up	GATTTTCTTTATTCGCCGTGAAGAGAATTTATG
EMSA CpG5_6 low	CATAAATTCTCTTCACGGCGAATAAAGAAA
EMSA CpG7 up	GCAGAAAACAGCCACGTGTTCCTGAAC
EMSA CpG7 low	GTTCAGGAACACGTGGCTGTTTTC
EMSA CpG7 mut up	GCAGAAAACAGCCATGTGTTCCTGAAC
EMSA CpG7 mut low	GTTCAGGAACACATGGCTGTTTTC
EMSA CpG8 up	AGCCCTATTCACCTTCGCTATGAGGGGGATC
EMSA CpG8 low	GATCCCCCTCATAGCGAAGGTGAATAGG
EMSA comp RAR	TCGAGGGTAGGGTTCACCGAAAGTTCACTCG
Biotin oligo ERE up	TTAATATTCCTGGTCAGCGTGACCGGAGC
Biotin oligo ERE low	TTAAGCTCCGGTCACGCTGACCAGGAATA
Biotin ΔG up	TTAATATTCCTGGTCACGTGACCGGAGC
Biotin ΔG low	TTAAGCTCCGGTCACGTGACCAGGAATA
Biotin CpG7 up	TTAAGAAAACAGCCACGTGTTCCTGAAC
Biotin CpG7 low	TTAAGTTCAGGAACACGTGGCTGTTTTC

Table S 2: EMSA and pull-down primers.

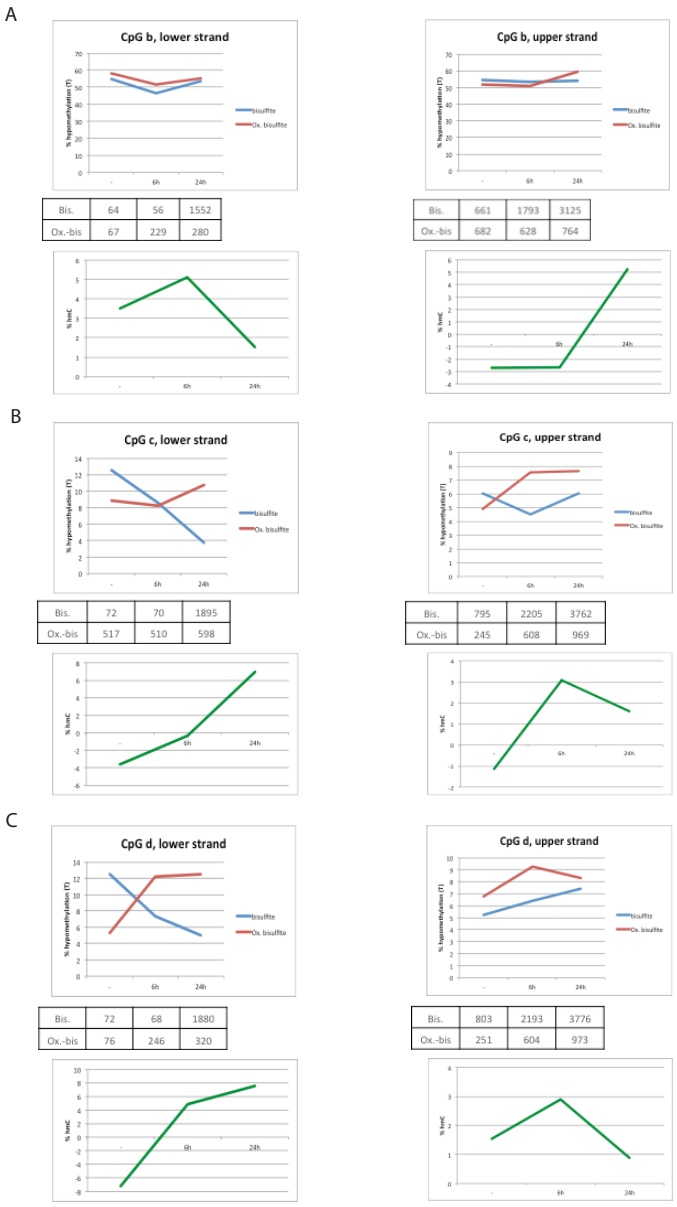
Mutagenesis primers	
new KpnI fwd	CCGGAGCTGAAAGAACACATGGTACCCGTGATTTCAATAAATAC
new KpnI rev	GTATTTATTGAAATCACGGGTACCATGTGTTCTTTCAGCTCCGG
destroy KpnI site fwd	CATGTTCACAACACTGGGGTATCTGATCTGTGGCTTCAGAG
destroy KpnI site rev	CTCTGAAGCCACAGATCAGATACCCCAGTGTTGTGAACATG
C57T fwd	GAAAGAACACATGGTACCTGTGATTTCAATAAATACATATG
C57T rev	CATATGTATTTATTGAAATCACAGGTACCATGTGTTCTTTC
C119T fwd	GCCTCTTAAATCATGTGTGTTGGTGCACATATGAATAC
C119T rev	GTATTCATATGTGCACCAACACACATGATTTAAGAGGC
C354T fwd	CCATGTACCAGCAGCCAGCTGTGACCCAATCTAGGAAAGC
C354T rev	GCTTTCCTAGATTGGGTCACAGCTGGCTGCTGGTACATGG
C410T fwd	CAATTTTAAATTTATTGTAAATGCTGTAGTAGAAGTGTTTTACTG
C410T rev	CAATTTTAAATTTATTGTAAATGCTGTAGTAGAAGTGTTTTACTG
C509T fwd	GTATTATTTGATTTTCTTTATTTGCCGTGAAGAGAATTTATGATTG
C509T rev	CAATCATAAATTCTCTTCACGGCAAATAAAGAAAATCAAATAATAC
C512T fwd	GATTTTCTTTATTCGCTGTGAAGAGAATTTATGATTGC
C512T rev	GCAATCATAAATTCTCTTCACAGCGAATAAAGAAAATC
C594T fwd	GAATTCAGCAGAAAACAGCCATGTGTTCCTGAACATTCTTCC
C594T rev	GGAAGAATGTTCAGGAACACATGGCTGTTTTCTGCTGAATTC
C655T fwd	GCAGAGCCCTATTCACCTTTGCCATGGAGGATGCCAAG
C655T rev	CTTGGCATCCTCCATGGCAAAGGTGAATAGGGCTCTGC
EREmut (C26T and C32T) fwd	GCTTAAAAATATTCCTGGTCAGTGTGACTGGAGCTGAAAGAACACA TGG
EREmut (C26T and C32T) rev	CCATGTGTTCTTTCAGCTCCAGTCACACTGACCAGGAATATTTTTAA GC
ΔG fwd	GCTTAAAAATATTCCTGGTCACGTGACCGGAGCTGAAAG
ΔG rev	CTTTCAGCTCCGGTCACGTGACCAGGAATATTTTTAAGC

Table S 3: Mutagenesis primers.

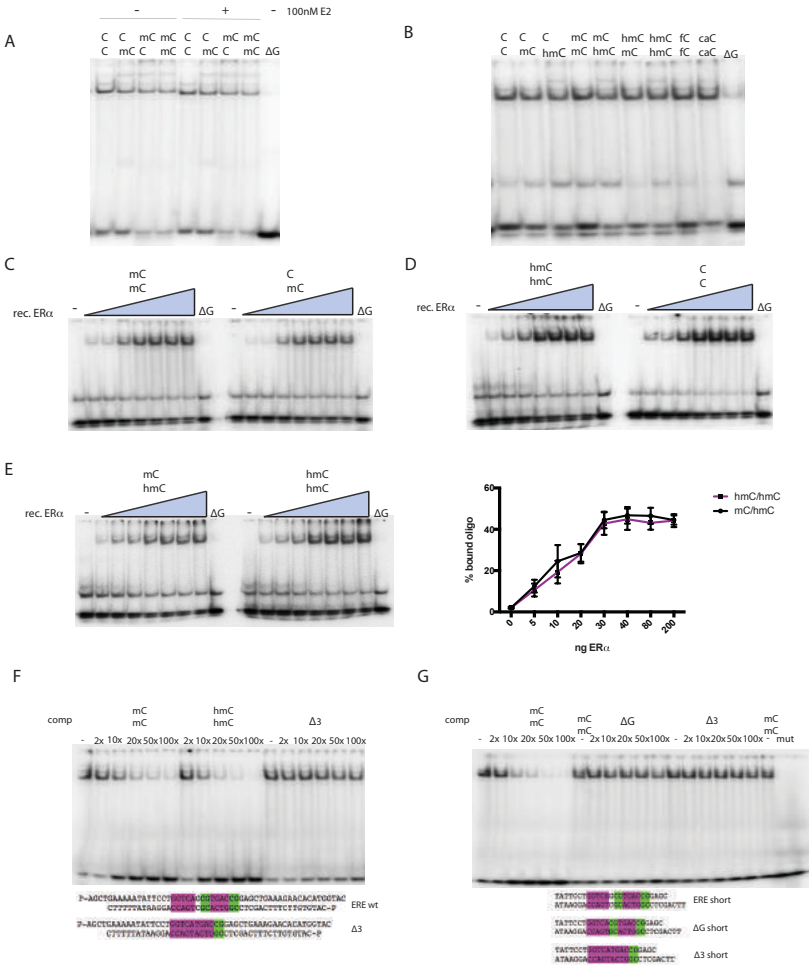
Kallenberger *et al.*, SI



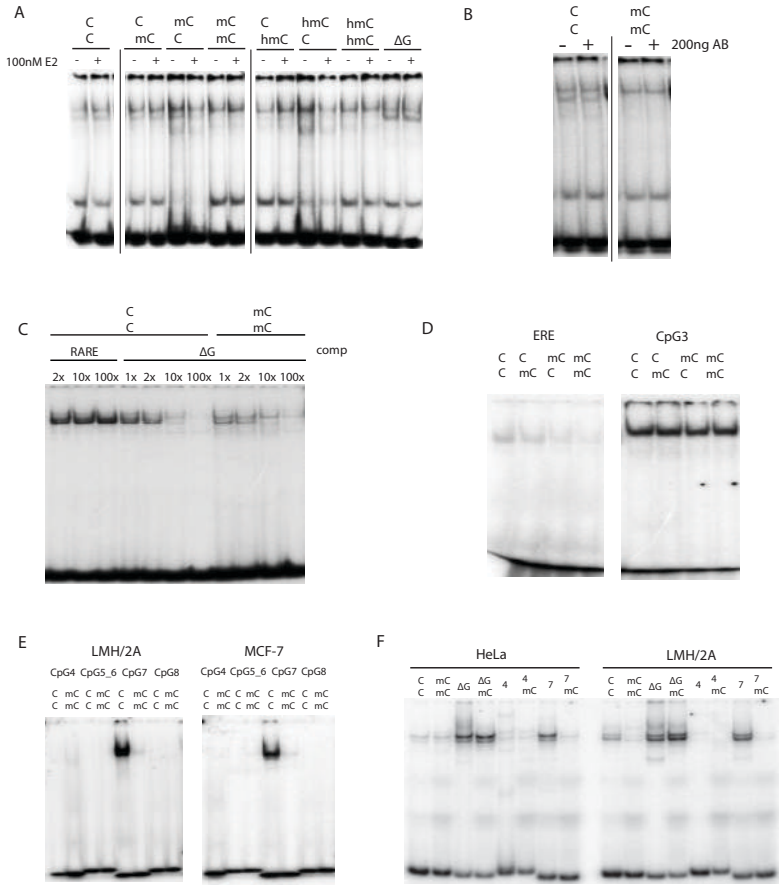
Supplementary Fig. 1: A Upper panel: mRNA levels were measured with qPCR after the indicated lenght of treatment with 100 nM E2 . One representative experiment is shown in technical triplicates. Data are represented as mean ± SD. Lower panel: Western Blot showing VTG protein levels and TFIIH as an internal loading control. **B** Relative ERα mRNA levels measured by qPCR after 6 and 72 h of E2 treatment. One representative experiment is shown in technical triplicates. Data are represented as mean ± SD. **C** Immunofluorescence of EdU incorporation without (left panels) or with 8 µM aphidicolin treatment (right panels) without (upper panels) and 24h (lower panels) of E2 treatment.



Supplementary Fig. 2: Bisulfite and oxidised-bisulfite sequencing result of the CpGs **b** (**A**), c (**B**) and **d** (**C**). The left panels show the results of the transcribed strand, the right panels show the non-transcribed strand. The numbers under the upper panels indicate the total amount of analyzed reads for every condition. The percentage of 5hmC (lower panels) was calculated as the difference of converted Cs between the oxidised bisulfite-converted DNA and only bisulfite-converted DNA.

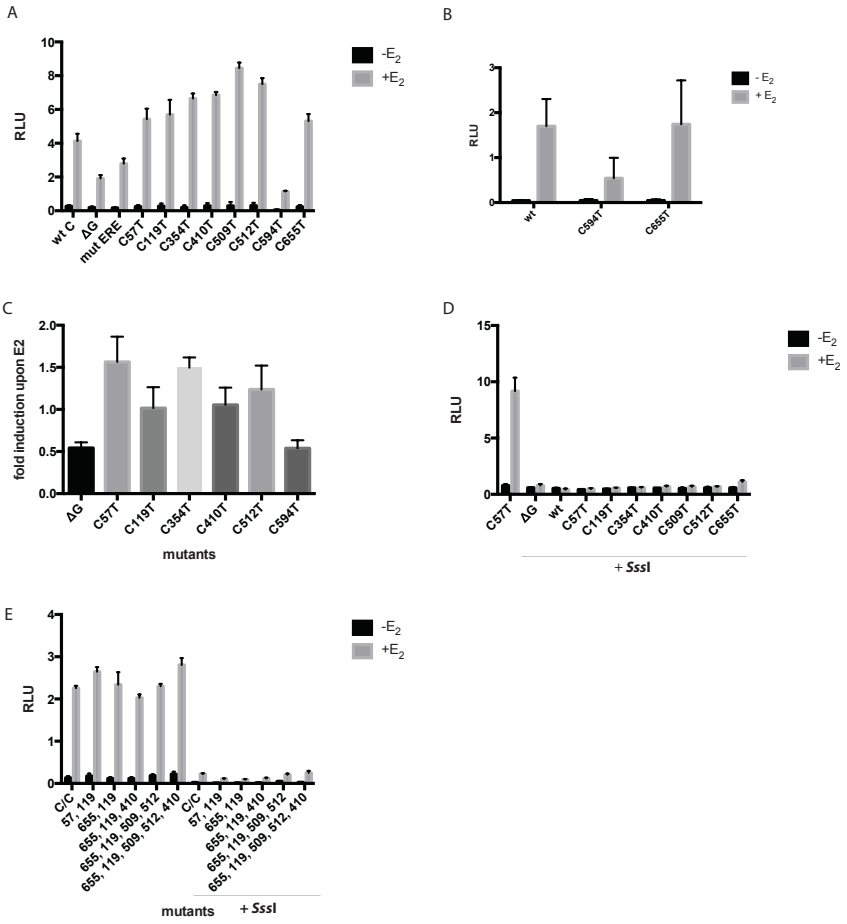


Supplementary Fig. 3: A EMSA showing binding of ERα to ERE and ΔG substrates of different methylation status, -/+ E2 treatment. **B** EMSA with different cytosine modifications in the ERE, including formylcytosine (fC) and carboxycytosine (caC). **C-E** EMSA with increasing concentrations of recombinant ERα (0, 5, 10, 20, 30, 40, 80 and 200 ng) and indicated cytosine modifications in ERE wt and ΔG substrates in the last lanes with 40 ng of ERα. Additionally in **E** right panel: Quantification of percentage bound oligo of three independent experiments of fully hemi- and fully-hydroxymethylated oligos. Data are represented as mean ± SD. **F & G** EMSA on unmethylated ERE with competition with increasing amounts of methylated or hydroxymethylated ERE , ΔG or Δ3 oligo (long oligos **F**, shorter oligos **G**).



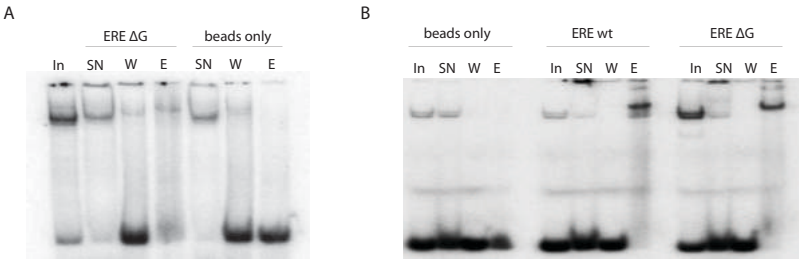
Supplementary Fig. 4: A EMSA with LMH/2A nuclear extracts and different combinations of modified oligos, -/+ E2. **B** EMSAs with addition of ERα antibody with unmethylated (left panel) and methylated (right panel) oligo in LMH/2A NEs. **C** EMSA in LMH/2A with increasing concentrations of a RARE consensus competitor or ΔG competitor on unmethylated (C/C) or methlyated (mC/mC) ERE. **D** EMSAs in LMH/2A NEs on ERE and CpG3 with combinations of different cytosine modifications showing complexes with distinct mobilities. **E** EMSA on CpG4, CpG5_6, CpG7 and CpG8 in LMH/2A (left panel) and MCF-7 (right panel) NEs, unmethylated and methylated. **F** EMSAs in HeLa (left panel) and LMH/2A (right panel) NEs on ERE, ΔG, CpG4 and CpG7, each unmethylated and methylated as indicated.

Kallenberger *et al.*, SI



Supplementary Fig. 5: A Luciferase assays of different mutants, unmethylated. Data are represented as mean of technical triplicates ± SD. **B** Luciferase assay of wt unmethylated plasmid (wt), C to T mutations in CpG7 (C594T) and in CpG8 (C655T). Data are represented as mean of three independent experiments ± SD. **C** Fold luciferase induction upon E2 treatment of different mutants. Data are represented as mean of three independent experiments ± SD. **D** Luciferase assays of different mutants, methylated with *Sss*I. Data are represented as mean of technical triplicates ± SD. **E** Luciferase assays of combined mutants, unmethylated and methylated with *Sss*I. Data are represented as mean of technical triplicates ± SD. experiments ± SD.

Kallenberger *et al.*, SI



Supplementary Fig. 6: EMSA on unmethylated ERE wt of proteins from affinity purification with the ERE wt or ΔG from HeLa (A) or LMH/2A (B) nuclear extracts. As a control, beads only eluents were used.

ADDITIONAL OBSERVATIONS

TET1 AND 2 AND MODIFIED CYTOSINES ARE PRESENT IN LMH/2A CELLS AND IN EMBRYONIC AND ADULT CHICKEN LIVER

TET1 and TET2 could be detected in the chicken hepatoma cell line LMH (Fig R1 A) and in embryonic and adult chicken liver. TET2 seems to be higher expressed in the adult tissue, whereas TET1 is more abundant in the embryo (Fig R1 C). Moreover, the presence of 5mC and the TET oxidation products could be confirmed using Dot Blots of LMH genomic DNA (Fig R1 B) and embryonic and adult chicken liver (Fig R1 D).

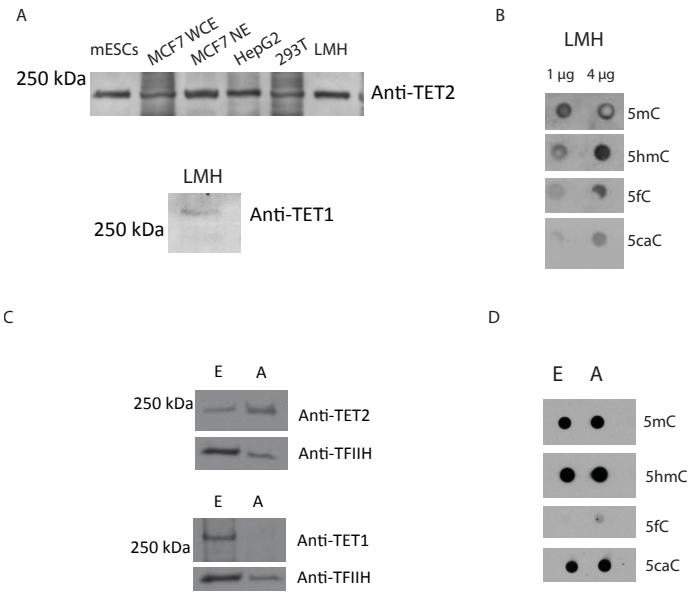


Figure R 1: Detection of TETs and cytosine species in LMH/2A and embryonic and adult chicken liver. A Western Blots of TET2 (upper panel) and TET1 (lower panel) in different cell lines. **B** Dot Blot of genomic DNA from LMH cells. **C** As in **A**, using whole cell extracts (WCE) from embryonic (E) and adult (A) chicken liver. TFIIH was used as the loading control. **D** As in **B**, but using genomic DNA extracted from embryonic and adult liver.

INVESTIGATION OF METHYLATION STATUS OF VITELLOGENIN ENHANCER AND VITELLOGENIN INDUCIBILITY IN CHICKEN EMBRYOS

We investigated the site-specific and inducible demethylation of the vitellogenin enhancer in chicken embryos, in addition to the LMH/2A cell line. To this end, we cut a "window" in the

shells of fertilized chicken eggs (Fig R2 A, [513]) and then added E_2 or EtOH from day 9 to day 10 of embryonic development. Interestingly, while CpGs **a**, **c** and **d** were ~90% methylated in the basal state, CpG **b** showed reduced methylation levels of 70-80% (Fig R2 B). This phenomenon is comparable to the adult rooster, where one strand of CpG **b** was unmethylated already prior to hormone treatment and the other strand 50% methylated, in contrast to 90% symmetrical methylation of the other CpGs [332]. Also, in LMH/2A cells, CpG **b** was a special case, being 50% methylated on both strands (Results II, Fig 1 B). We could show a weak induction of *VTG* expression in 10-day embryos after 24 h of E_2 treatment compared to EtOH mock treatment (Fig R2 C), but no detectable demethylation, similar to that seen in the hepatoma cell line LMH/2A.

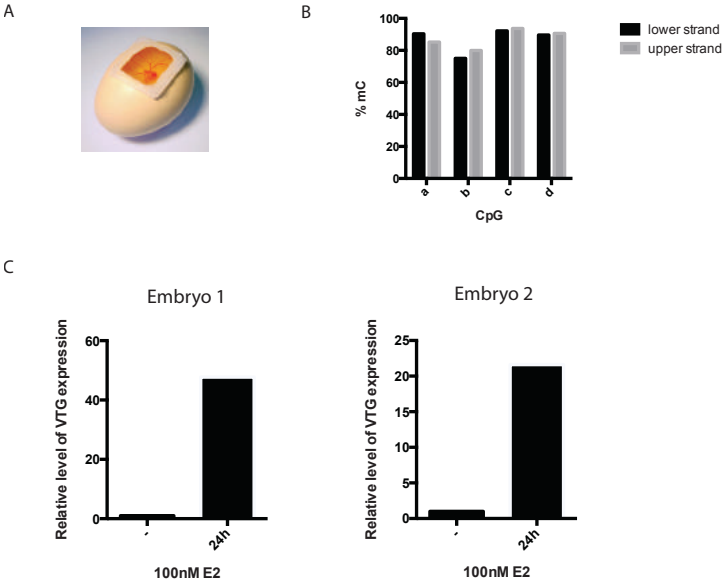


Figure R 2: Manipulations of chicken embryos *in ovo*. **A** Eggs were 'windowed' to enable access for manipulations. **B** Basal methylation states of the four CpGs described by Saluz *et al.*, assessed by PacBio single molecule sequencing. **C** Induction of vitellogenin expression in two individual embryos.

MASS SPECTROMETRIC IDENTIFICATION OF PROTEINS BINDING THE WT AND ΔG VITELLOGENIN ERE

As we saw a mobility shift in LMH/2A-, MCF-7- and HeLa NEs that was methylation-sensitive and distinct from ERα, we set out to identify the proteins binding the ERE wt and ΔG sequence using an approach that combined affinity-enrichment and mass spectrometry (MS). To improve the yield, we generated multimers of the oligonucleotides by end-to-end ligation, filled-in the overhangs with bio-dUMP and attached the DNA to Streptavidin-derivatised Dynabeads. Mass

spectrometric analysis of the proteins bound to these substrates upon incubation with HeLa extracts identified several proteins binding the ERE that were absent in the control (Table 1). A candidate that caught our attention was Max-like protein X (MLX). MLX is a helix-loop-helix transcription factor, known to bind the E-box sequence CTCGAG. The deletion of the G in the ERE spacer in ΔG generates an E-box motif and CpG7, shown to bind the same protein complex in NEs, is located in an E-box (Results II).

Table 1: Proteins enriched from HeLa NE. Proteins of the MAX network are marked in green.

Gene	MW (kDa)	Spectrum ΔG	Spectrum beads	Probability if detected
MLXPL	93	19		>95%
MLX	33	13		>95%
NR2F6	43	11		>95%
MSH6	153	18	4	>95%
NR2C2	65	20		>95%
RXRB	57	10		>95%
MSH2	105	12	6	>95%
UBP1	60	15		>95%
MPG	33	12		>95%
MLXIP	101	19		>95%
ESRRA	46	9		>95%
ZBTB9	51	9		>95%
ZBED1	78	10		>95%
RFX5	65	8		>95%
NR2C1	67	9		>95%
MGA	332	7		>95%
MAX	18	4		>95%

We performed the same pull-down/ MS experiment also with LMH/2A NEs and could also retrieve MLX, specifically in the affinity purification with ERE wt and ΔG, but not in the control (Table 2). We could also detect ERα in the enrichment from LMH/2A NEs, confirming the specificity of the method.

Table 2: Proteins enriched from LMH/2A NE. Proteins of the MAX network are marked in green.

Gene	MW (kDa)	Spectrum ERE	Spectrum ΔG	Spectrum beads	Probability if detected
Gga	147	64	63		>95%
SALL4	118	56	58		>95%
RXRG	51	40	35	2	>95%
CHD4	219	35	31		>95%
KIA1958	79	20	20		>95%
ESR1	67	16	11		>95%
HNF1A	69	13	8		>95%
CREBBP	265	13	13		>95%
RARG	49	8	11		>95%
NR2C1	62	9	8		>95%
MPG	30	5	6		>95%
NR2F1	46	6	7		>95%
MSH2	95	6			>95%
TBX3	78	4	4		>95%
HNF4A	51	5	5		>95%
CTBP1	47	5	7		>95%
MSH6	142	6	4		>95%
TDG	45	4	6		>95%
LXRA	47	3	4		>95%
NR2C2	65	5	4		>95%
MLXIP	89		6		>95%
ZBTB17	89	3	5		>95%
MLX	28		4		>95%
MAX	16		2		>95%

To verify that MLX is involved in the complex, causing the mobility shift with ERE wt, ΔG and CpG7 substrates, we performed EMSA-supershift experiments, using a monoclonal antibody against MLX and other proteins identified in the MS analysis. Only the MLX antibody caused a supershift (Fig R3), indicating the presence of MLX in the complex that formed the mobility shift on the substrate. The supershift was, however, only detectable on the ΔG and not on the CpG7 substrate. This unanticipated finding thus needs further study.

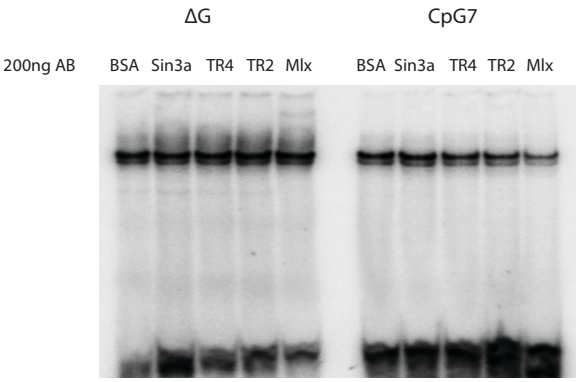


Figure R 3: Supershift experiments show evidence for the role of MLX in factor X shift. 200ng of the indicated antibodies were added to the EMSA reaction with LMH/2A NE. The mobility shifts were carried out with the ΔG- (left panel) or the CpG7 substrate (right panel).

LOCALIZATION AND ACTIVITY OF EXOGENOUS TEN-ELEVEN TRANSLOCATION PROTEINS IN HEK 293 CELLS

We checked the localization of human TET1 and mouse TET2 full-length proteins upon expression in HEK293 cells. Both proteins were expressed in the cells 24 h after transfection, as assessed by immunofluorescence against their Flag tag (Fig R4 A). They were successfully targeted to the nucleus and colocalized with DAPI staining. We also analyzed the DNA of transfected cells to assess whether the amount of the oxidized base 5hmC was increased. To this end, we performed an adapted fibre assay, where we stained DNA fibres with 5mC and 5hmC antibodies, followed by staining with fluorescent secondary antibodies. 5hmC was not detectable in mock transfected cells, but clearly enhanced in the cells transfected with full length human TET1. The 5hmC signal displayed nicely on the fibres, similar to 5mC (Fig R4 B).

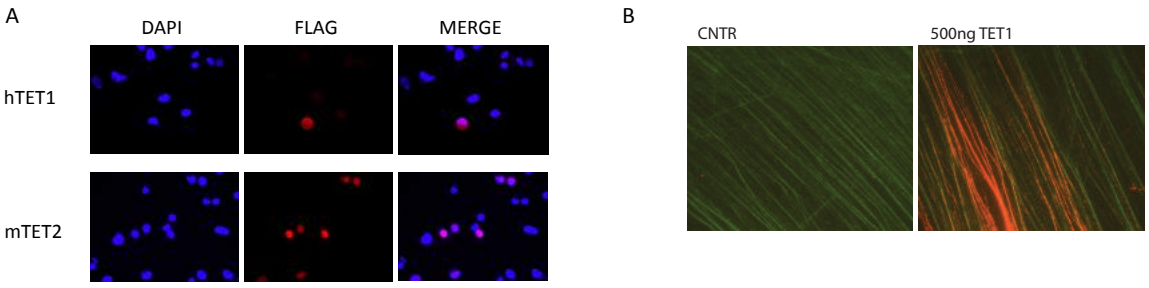


Figure R 4: A Immunofluorescence using DAPI and Anti-Flag antibody in HEK293 cells transfected with full-length human TET1 or mouse TET2. **B** DNA fibres stained with Anti-

5mC and Alexa Fluor 488 and Anti-5hmC and Alexa Fluor 568 antibodies from DNA of mock (CNTR) or human TET1-transfected HEK293 cells.

TET1 CATALYTIC DOMAIN PURIFICATION AND ENZYMATIC ACTIVITY TESTS

We purified active human TET1-CD from stably-expressing HEK293 FlpIn cells using a 3x Flag tag, as described in 'Materials and Methods of Additional observations'. Input, flowthrough (FT) and elutions (E1 and E2) were loaded on a 6% polyacrylamide SDS gel and stained with Coomassie blue to verify expression and successful enrichment with the Flag beads (Fig R5 A). The TET1-CD was obtained in high purity. The faster-migrating bands on the gel most likely represent degradation products and the bands that migrated slower might be oxidized TET1-CD products, as observed by others [28]. Activity of the purified protein was assessed by carrying out a TET reaction on 5mC bearing substrates, as described in 'Materials and Methods of Additional observations'. The small-scale- (26.8.15) and the large-scale (02.10.15) purifications from 100 15 cm dishes were active on a 5mC-containing plasmid, as assessed by Dot blot using a 5hmC specific antibody (Fig R5 B).

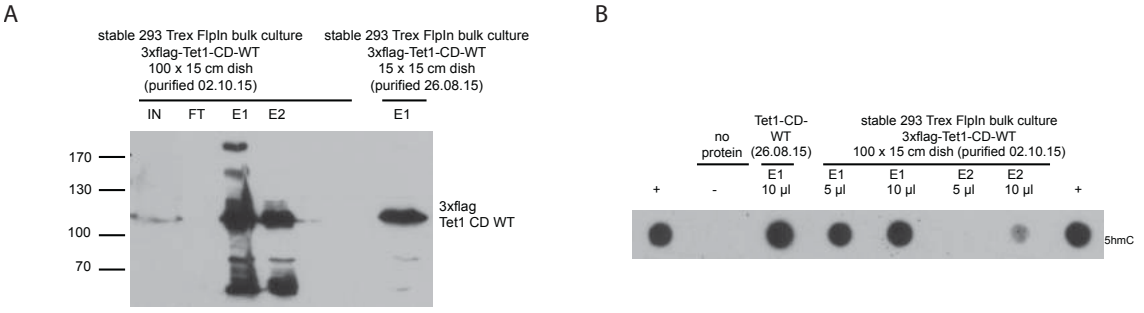


Figure R 5: A Coomassie blue-stained SDS-PAGE of purified TET1-CD small-scale (26.08.15) and large-scale (02.10.15) preparations. Input (IN), flow-through (FT), elution 1 (E1) and elution 2 (E2) of the large scale and E1 of the small scale were loaded onto a 6% gel**. B** Activity test of purified proteins on 5mC-containing plasmid DNA. Product DNA was loaded onto a Dot blot and stained with Anti-5hmC antibody. Lanes marked "+" represent control 5hmC-containing plasmid DNA.

We next addressed the question, at which pH the TET reaction was most efficient, as there were different conditions reported. We performed the TET reaction at pH 7.9 and at pH 6.5 using our purified TET1-CD and oligonucleotide substrates. A Dot blot assay revealed that 5mC was converted more efficiently to 5hmC at pH 6.5 (Fig R6 A). 5hmC-containing oligonucleotides were

used as a positive control. Also the subsequent oxidation steps to 5caC were more efficient at pH 6.5 (Fig R6 B).

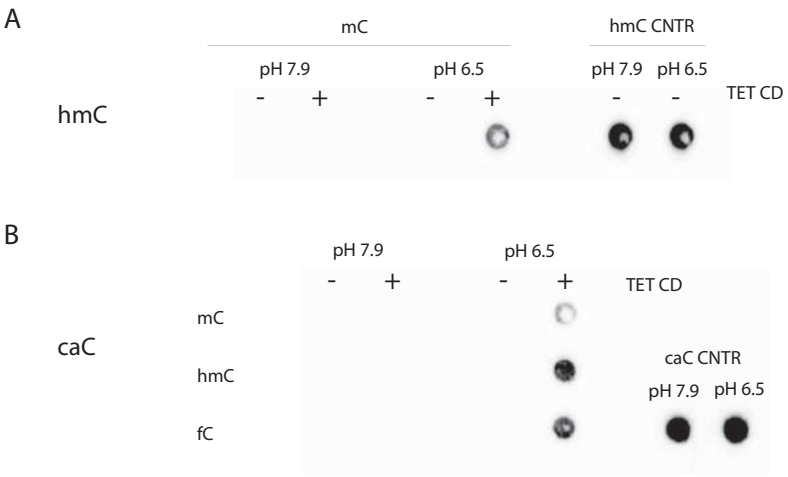


Figure R 6: TET reactions using the TET1-CD from the large-scale purification in Fig R5 at different pH. A Starting from 5mC-bearing oligonucleotides, reaction products were visualized with Anti-5hmC antibodies and HRP-coupled secondary antibodies. 5hmC-containing oligos were used as positive controls (hmC CNTR). **B** Starting from 5mC (mC), 5hmC (hmC) or 5fC (fC) oligos, product formation was assessed by staining with an Anti-5caC antibody. 5caC-containing oligos were used as positive controls (caC CNTR).

To address some of the open questions concerning TET processivity, we plan to use our purified active TET1-CD and the established reaction conditions. Studies performed so far only assessed average product formation of the TET proteins, either *in vivo* in genome-wide studies or biochemically. No information concerning single oxidation steps at defined loci can be retrieved from these studies. We established an approach on covalently-closed oligonucleotide substrates, sequenced with the PacBio technology that provides information on single molecules. This enables us to assess whether the TET reaction yields asynchronous oxidations or whether they primarily oxidize both strands symmetrically. As the substrate contained two CpGs, we are also able to get information about processivity along the DNA on each strand (Fig R7).

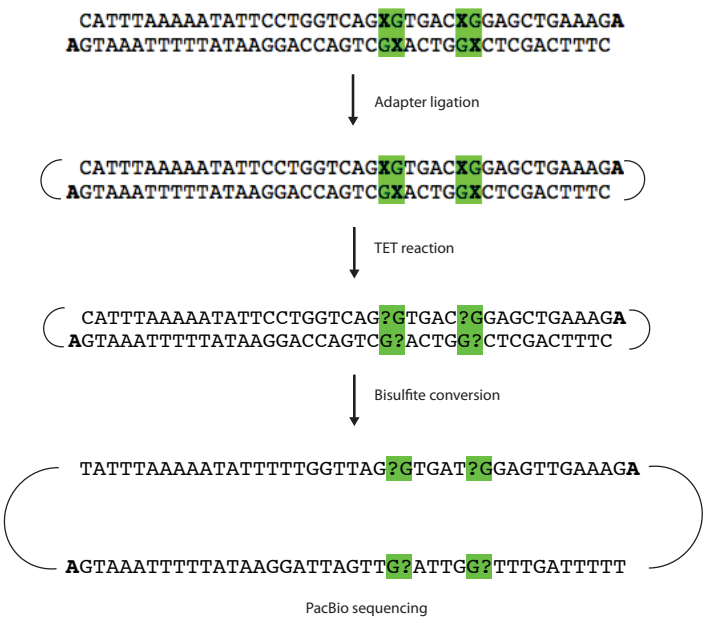


Figure R 7: Schematic representation of the work-flow for the biochemical characterization of the TET reaction.

In preliminary experiments, we obtained few reads when starting with short oligonucleotides, confirming that the sequencing primer for the bisulfite converted adapter works. The loss during the purification steps was however big and the procedure needs further improvement. In additon, the optimal TET reaction incubation times and protein concentrations have yet to be optimised, so as to enable us to see single oxidation steps.

MATERIALS AND METHODS TO 'ADDITIONAL OBSERVATIONS'

Antibodies

Name	Origin	Provider	Dilution
Anti-TFIIH	rabbit	Santa Cruz	1:1000
Anti-TET1	rabbit	Abcam	1:1000
Anti-TET2	mouse	Abiocode	1:1000
Anti-5mC	mouse	Serotec	1:500
Anti-5hmC	rabbit	Active motif	1:10'000
Anti-5fC	rabbit	Active motif	1:2000
Anti-5caC	rabbit	Active motif	1:2000
Anti-rabbit Alexa 568	goat	Molecular Probes	1:300

Anti-mouse Alexa 488	goat	Molecular Probes	1:300
Anti-flag	mouse	Sigma	1:1000
Anti-MLX	mouse	Santa Cruz	1:1000
Anti-Sin3a	mouse	Santa Cruz	1:1000
Anti-TR2	mouse	Santa Cruz	1:1000
Anti-TR4	mouse	Santa Cruz	1:1000
Anti-rabbit IgG HRP	goat	GE Healthcare UK	1:5000
Anti-mouse IgG HRP	sheep	GE Healthcare UK	1:5000

Cell culture

Human embryonic kidney (HEK) 293 cells were grown in DMEM medium (Gibco), supplemented with 10% FCS and streptomycin/penicillin (100 U/ml). Cells were grown at 37°C in a 6% CO_2 humidified atmosphere.

Dot Blot

Samples were heated for 10 min at 100°C in a total volume of 250 µl in 0.4 M NaOH, 10 mM EDTA and transferred quickly to ice to avoid re-annealing. An equal volume of 2 M ice-cold ammonium acetate was added. After pre-soaking in 1 M ammonium acetate, the nylon membrane was inserted into the Dot Blot apparatus according to manufacturer's instructions. Wells were rinsed with water before applying the samples. The vacuum was kept low to avoid deforming the membrane. After samples were blotted, the wells were washed once with 1 M ammonium acetate, the membrane was removed and soaked for 5 min in 5x saline-sodium citrate (SSC; 0.75 M NaCl, 75 mM sodium citrate pH 7). The DNA was crosslinked to the membrane by either baking at 80°C for 2 h (plasmids) or by irradiating the membrane with 120 mJ UV irradiation (for oligos; Stratalinker 1800, Stratagene). The membrane was rinsed with 5x SSC, followed by blocking in 5% milk/TBS-T (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% Tween-20) for 1 h. Incubation with the primary antibody was performed overnight at 4°C. The membrane was washed 3x for 5 min in TBS-T, incubated with the secondary antibody for 1 h at RT and washed again, before being developed in a Fusion Solo (Vilber Lourmat) using WesternBright Chemiluminescent Detection Reagent.

Oligos used for activity tests: 5'-CCAGACGTCTGTCGACGTTGGGAAGCTTGAG-3', with the underlined cytosine carrying the modification. Plasmids used for activity tests were prepared by primer extensions with the above oligos on pRichi 350t SalI phagemid ssDNA [514].

Manipulation of chicken embryos

Eggs were prepared as described [515, 516]. At day 10, eggs were treated with 100 nM E2 by pipetting ~55 µl (depending on weight of egg) of a 100 µM solution (10% EtOH) into the egg, through the window prepared at day 3 of development at 39°C. 24 h later, the eggs were broken and the dissected out of the embryo for RNA and protein extraction. Livers were homogenized using a Tissue Ruptor (Qiagen) in RIPA buffer for proteins and in RLT buffer for RNA. Following homogenization, protein samples were processed further according to the Western Blot protocol as described in Results II and RNA was extracted using an RNAeasy Kit (Qiagen).

Immunofluorescence

Cells were grown on coverslips in 6-well plates and transfected with plasmids (mTET2; FH-Tet2-pEF, a gift from Anjana Rao (Addgene plasmid # 41710)[446] and 3x Flag hTet1-ORF using Lipofectamine 2000 according to manufacturer's instructions. Coverslips were washed in phosphate-buffered saline (PBS) and cells were fixed in 3.7% formaldehyde/PBS for 15 min. Coverslips were washed 3x in PBS before cells were permeabilized in 0.2% TritonX-100/PBS for 5 min. Coverslips were washed 3x in PBS. 50 µl of 3% milk/PBS were pipetted onto Parafilm and the coverslips were put cells-down onto the mix and blocked for 15 min. The same procedure was applied with the primary antibody dilution in 3% milk/PBS and coverslips were incubated at 4°C overnight. The coverslips were washed 3x in PBS and incubated with the secondary, fluorescent antibody in 3% milk/PBS at 37°C for 1 h, shielded from light. After 3 washing steps in PBS, cells were fixed with mounting media containing 4',6-diamidino-2-phenylindole (DAPI, VectaShield). Slides were analyzed on an Olympus IX81 fluorescence microscope.

DNA Fibre staining

Cells were washed and scraped off in PBS. 2 µl of a cell suspension, containing approximately 1x 10^6 cells/ml, were pipetted onto a glass slide (Super Frost Plus, Menzel-Glaser). 7 µl of lysis buffer (200 mM Tris-HCl pH 7.4, 50 mM EDTA, 0.5% SDS) were added, cells were mixed with a pipette tip and incubated for 2 min at RT before glass slides were slightly tilted to let the drops run down. Once slides were dry, DNA was fixed in methanol/acetic acid 3:1 for 10 min at RT. Before staining, slides were washed twice in PBS and DNA was denatured in 2.5 M HCl for 1 h. Slides were washed 5x in PBS for 3 min each and blocked in 2% bovine serum albumin (BSA) and 0.1% Tween-20 in PBS for 45 min. Slides were incubated in primary antibody for 2.5 h at RT, washed 5x in PBS with 0.2% Tween-20 (PBS-T) and once in blocking solution. Incubation with the secondary, fluorescent antibody was carried out at RT for 1 h. Slides were washed again 5x with PBS-T, before cells were fixed with mounting media containing DAPI (VectaShield). Slides were analyzed on an Olympus IX81 fluorescence microscope.

Preparation of oligos for enrichment of proteins for MS

Oligos with compatible 5'-overhangs were annealed in annealing buffer (50 mM Hepes, pH 7.5, 100 mM NaCl) to 10 µM (ERE wt upper 5'-TTAATATTCCTGGTCAGCGTGACCGGAGC-3'; ERE wt lower 5'-TTAAGCTCCGGTCACGCTGACCAGGAATA-3'; ERE ΔG upper 5'-TTAATATTCCTGGTCACGTGACCGGAGC-3'; ERE ΔG lower 5'-TTAAGCTCCGGTCACGTGACCAGGAATA-3'). 400 pmoles of annealed ds oligos were ligated end-to-end with 10 U T4 ligase (NEB) overnight at 16°C. Efficient ligation was verified on an agarose gel and the DNA was recovered by EtOH precipitation. Overhangs were filled-in according to the protocol for 3'-labeling, but with only 1 mM biotin-11-dUTP (ThermoScientific) and 2 µl [α-^{32}P]-dATP in 40 µl. Reaction was incubated for 10 min at RT before 1 µl of 10 mM cold dATP was added for another 10 min. Mixture was passed twice through a Sephadex desalting column to get rid of the free biotin.

Enrichment of binding proteins for MS

30 µl of Dynabeads M-280 Streptavidin (ThermoFisher) per oligo were washed with 1 ml of washing buffer (5 mM Tris, pH 7.5, 0.5 mM EDTA, 1 M NaCl). The beads were taken up in 30 µl of washing buffer. The ligated oligos from above were added and allowed to bind to the beads, until the beads were saturated, assessed by the lack of radioactive signal in the flow-through. The beads were washed once and taken up in 2x EMSA binding buffer without BSA (10 mM Tris-HCl, pH 7.6, 50 mM NaCl, 1 mM DTT, 5% glycerol). 1.5 mg of nuclear extract were pre-incubated with 30 µg of poly(dI.dC), before an equal volume was added to the beads (split into three tubes). The mixture was incubated for 15 min with gentle shaking at 4°C. Supernatant was collected and beads were washed with 50 µl of 1x EMSA binding buffer without BSA. Wash fraction was collected and proteins were eluted in 50 µl (for all three tubes) of elution buffer (10 mM Tris-HCl, pH 7.6, 500 mM NaCl, 5% glycerol) for 10 min, shaking at 4°C. 7 µl of the eluted fraction were used for EMSA to verify the presence of the binding proteins.

Shotgun LC-MS/MS

The protein mixture was digested with trypsin into peptides. After cleaning up, the peptide mixture was applied to a reversed-phase high-performance liquid chromatography (HPLC) column and separated prior to ionization and analysis by the mass spectrometer. The peptide ions, selected by the instrument, were fragmented and the resulting fragment ion spectrum was recorded. It provided a set of peptide signatures, which were used to identify the respective polypeptides by the database search algorithm MASCOT. The factors were scored according to the average probability using total spectra counts.

Purification of TET1-CD from HEK293 cells stably expressing TET1-CD wt

Stable cell lines were established using 293 TrexFlpIn cells that were transfected with 200 ng of pTO-pAIO-FRT-3x Flag Tet1CD WT (from aa 1418-2136) containing hygromycin resistance and 1.5 µg of pOG44 Flp-Recombinase expression vector (ThermoFisher). When confluent, cells were plated in selection medium containing Hygromycin (250µg/ml, InvivoGen) and Blasticidin (10µg/ml , InvivoGen). Once visible colonies formed, they were trypsinized and re-seeded as bulk culture.

Approximately $1.2x10^9$ of 293 TrexFlpIn pTO-pAIO-FRT-3x Flag Tet1CD WT bulk culture were treated with 100 ng/ml Doxycyclin (Clontech) for 24 h, before being harvested at around 50% confluency directly in PBS on ice. Cells were spun for 5 min at 500 x g at 4°C, washed in cold PBS and spun again. Pellets could be snap frozen and stored at -80°C for later purification.

Pellet was resuspended in 12.5 ml cold lysis buffer (50 mM Tris pH 7.5, 1 mM EDTA, 1% TritonX-100, 300 mM NaCl, 1 mM PMSF, 2x protease inhibitor cocktail, Roche) and incubated 20 min on ice. Lysate was spun for 10 min at 14'000 rpm at 4°C, supernatant was transferred to a new tube and diluted 1:1 in 50 mM Tris pH 7.5, 1 mM EDTA, 1% TritonX-100, 1 mM PMSF, 2x protease inhibitor cocktail. An Input sample for the SDS gel was kept aside.

For the immunopurification (IP), 500 µl Anti-Flag M2 beads (Sigma) were equilibrated by washing 3x in 1 ml of washing buffer (50 mM Tris pH 7.5, 150 mM NaCl) and spinning for 2 min at 500 x g. All steps were carried out at 4°C. The lysate was added to the beads and incubated on a rotating wheel for 3 h. A flow-through (FT) sample for the SDS gel was kept aside. The beads were washed 3x in 1 ml washing buffer by incubating on the rotating wheel for 5 min, followed by spinning for 2 min at 500 x g. Proteins were eluted in two rounds (E1 and E2) with 500 µl elution buffer (50 mM Tris pH 7.5, 50 mM NaCl, 1 mM β-mercaptoethanol, 5% glycerol, containing 200 ng/ml Flag peptide (3x Flag peptide, Sigma) by incubating them for each round for 30 min on a rotator. A sample was kept aside for the SDS gel and the rest was aliquoted and snap frozen and stored at -80°C.

TET enzymatic reaction

For the TET reaction, 10 µl of the purified protein (E1) were used for 500 ng of 5mC-containing plasmid in a volume of 15 µl in TET buffer (prepared freshly as 10x: 20 mM ascorbic acid, 10 mM DTT, 1 mM ammonium iron(II) sulfate (FAS), 10 mM 2-oxoglutarate, 500 mM NaCl, 10 mM ATP, 500 mM Tris pH 6.5 or 7.9, as mentioned in 'Additional observations', Fig R6). Reaction was incubated for 3 h at 37°C and heat inactivated for 5 min at 85°C.

Adapter ligation for PacBio sequencing of short oligonucleotides

1 µg of ds oligonucleotides were mixed with 1 µl of 20 µM overhang adapter (Pacific Biosciences). Provided template preparation buffer was added and supplemented with 1.5 µl of 1 mM ATP. 30 units of provided ligase (Pacific Biosciences) was added, reaction was made up to 30 µl and incubated for 2 h at 25°C. After heat inactivation at 65°C for 10 min, reaction was treated with exonucleases III (50 U) and VII (5 U) to get rid of unligated oligos and adapters. Reaction mixture was incubated at 37°C for 1 h. Reaction was purified on AMPure®PB Beads (Pacific Biosciences) according to manufacturer's instructions. Ligated oligos were now ready for TET reactions and bisulfite conversion, before being sequenced using the PacBio technology. As the adapters provided by the company also undergo bisulfite conversion, a different sequencing primer had to be added (5'-TTTTTTTTTTTTTTTTTTTCTCTCAACAACAACAAC-3', the underlined nucleotides containing 2'-methoxy modifications to increase stability).

Western blots, DNA- and RNA extractions, PacBio sequencing, RT-qPCR and EMSAs were carried out as described in Results II.

DISCUSSION

DSB INDUCTION RESULTING FROM AN INTERPLAY BETWEEN BER AND MMR

The experimental set-up developed in our lab, using defined, lesion-containing plasmid substrates and nuclear cell extracts, represents a very nice system to study the DSB-induction step in class-switch recombination. The plasmids do not replicate, mimicking the situation in germinal center B-cells that are in the G1-phase of the cell cycle. The defined locations of the lesions enable us to draw conclusions about the mechanism of action of the cellular repair enzymes acting on them.

Based on our *in vitro* and *in vivo* experiments (Results I, [517]), we propose a model for the combined action of BER and MMR in enabling efficient CSR (Results I and Fig D1). U/G mispairs are substrates for both BER and MMR. This leads to a competition of the pathways for the lesions, whose outcome depends on lesion-density, sequence context and cell cycle phase. We propose that DSBs occur through a collision of repair intermediates of BER and MMR, ss nicks and ss degradation tracts, which individually would not form DSBs, but - when combined - very efficiently do so.

We see an important contribution of MMR, namely MutSα, in the induction of DSBs in our *in vitro* and *in vivo* systems, and in the generation of mutations *in vivo*. The break formation is completely dependent on the BER enzyme UNG, on the other hand. An explanation for these differences might be that MMR alone lacks a strand-discrimination signal, necessary for efficient repair. Is has already been shown that MMR can make use of sites of ongoing base excision repair [118] and sites of ribonucleotide excision [129] to start EXO1-mediated degradation of one strand. It is therefore conceivable that CSR is highly dependent on these UNG-created nicks. This phenomenon has also been observed in *Ung*$^{-/-}$ mice [64]. The milder effect on DSB formation we see in the absence of MMR is also consistent with genetic studies [137, 139, 245]. One explanation could be that repair tracts generated after excision of uracils by UNG could result in DSBs in case they are located in close proximity on opposite strands. In these situations, MMR degradation tracts would not be needed. We showed in our system, however, that nicks on opposite strands have to be very closely spaced to result in melting of the ds DNA between them. A stretch of 50 bp between the uracils is enough to keep the DNA stably linked. It is imaginable that the sequence context in certain parts of the S regions provides many AID target sites that could be processed into MMR-independent DSB, whereas other regions might depend on MMR to allow formation of DSBs. This could account for the differences in Ig isoform distributions when MMR is absent.

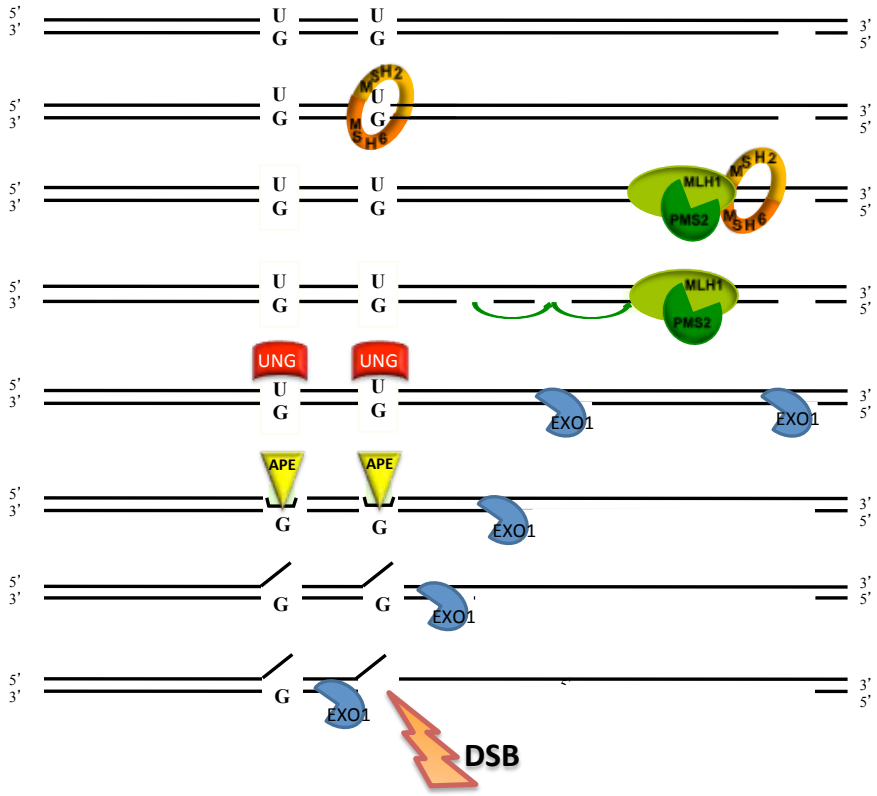


Figure D 1: Model of the interplay of BER and MMR in the processing of uracil lesions in DNA. In the case when the MMR system is faster than BER in recognizing and binding a uracil, MutSα is activated and recruits MutLα that sets additional nicks into the already-nicked strand. EXO1 is loaded at the nicks and starts degrading one strand with its 5' to 3' polarity. In the meantime, other uracils generated in the vicinity might have been addressed by the fast UNG and APE1, resulting in additional nicks. If the EXO1 degradation tract runs into the BER created lesions, DSBs are the consequence. Adapted from [517].

The finding that the absence of MMR in our system had such a great impact was somewhat surprising, as the absence of MMR was reported to result in a less severe phenotype than BER-deficiency in mice [64, 137, 139, 245], as mentioned above. By providing a nick in the substrates, we do, however, stress the involvement of MMR, as the nick can be used as a strand-discrimination signal and initiation site for the EXO1-mediated degradation tract. A more physiologically-relevant substrate would contain uracils on both strands, instead of a nick on one strand. These substrates were extremely complex to prepare and required carefully-designed purification steps. Preliminary results obtained with those plasmids gave puzzling results. We completely lost MMR-dependence in break formation using this substrate. We did see efficient break formation, however, dependent on UNG and on catalytically-active EXO1. These results support the notion that BER, initiated by UNG is extremely fast and efficient. This could mean

that MMR does not have access to the uracils under normal circumstances. However, MMR is clearly involved in CSR. An explanation for this could be that during AID-induced CSR, there are so many uracils created in close proximity that the capacity of BER is exceeded, giving MMR the chance to reach some of the lesions.

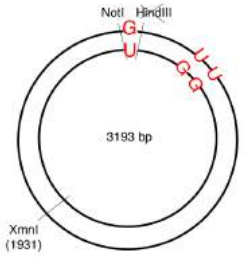


Figure D 2: Schematic drawing of the CSR-substrate bearing uracils of both strands.

The question why we still see break formation remains. The uracils are too far apart to result in melting of the DNA strands. This is confirmed by the EXO1 dependence. The obvious conclusion to draw here is that EXO1 is loaded at the nicks resulting from BER independently of MMR, thereby assisting in break formation. Why this is happening only in the substrate containing uracils on both strands (Fig D2) and not in the substrate with the nick investigated in Results I cannot be answered conclusively here. The only difference between the substrates is the nick versus the uracil at a single site. Maybe the physical presence of the BER or the MMR machinery at the nick is needed for loading of EXO1 and their targeting might be inefficient in the case of a pre-existing nick, as compared to a uracil.

MUTATIONS VERSUS DBS: SHM VERSUS CSR

An interesting question concerning the mechanisms of SMH and CSR is how the outcomes of lesion processing can be so different, considering that the initial steps are equal. Both pathways are initiated by AID-induced cytosine deaminations, creating uracils in DNA. In SHM, these lesions are converted into mutational hotspots in the variable regions, whereas in CSR the results are targeted DSBs involving loss of the intervening DNA.

One important point to keep in mind is that the outcomes might be less distinct from each other than might appear to be the case in B-cells. We do see a considerable amount of mutations in our CSR system, highly dependent on MMR (Results I), similar to SHM. Although cytosines in WRC AID targeting sites were published to be mutational hotspots [215], we do not see any accumulation of mutations at our uracil sites above random levels. It is therefore conceivable

that the same non-canonical, highly-mutagenic MMR mechanism acting in SHM is also responsible for the mutations we detect in our substrates [117, 148, 149].

DSB formation, on the other hand, does also occur during SHM, mostly resulting in B-cell death and therefore not easily measurable. Another point to keep in mind is that DSB formation is generally an infrequent even, not only in SHM, but also in CSR. The power of the process in B-cells lies in the efficient selection and elimination steps. In our *in vitro* system that was optimized to assess DSBs by having a nick on the opposite strand to the uracils, DSB incidence never exceeded 30%, reached by addition of recombinant proteins. In the *in vivo* system, we had to introduce two selection steps in order to be able to detect the DSB events. Moreover, we assume that many DSBs escaped our analysis, because they did not lead to efficient re-annealing. In B-cells, these DSBs, possibly also happening during SHM, would lead to apoptosis. Higher eukaryotes can easily tolerate high apoptosis rates in B-cells, compared to other tissues.

Nevertheless, the balance of DSBs versus mutations is clearly different in SHM versus CSR. As both processes occur in germinal center B-cells and therefore in the same environments, the differences have to come from the exact sequence contexts and the targeting of AID to these sites, as well as from the re-annealing step after DSB induction. Supporting this notion, it was observed that the sequence context of the Sμ region is critical for CSR [518, 519]. Another observation made was that transcription targeted to S regions results in the formation of R-loop structures, whereas these structures are inefficiently formed in the variable regions [116].

Further investigations of the differences in sequence context in S regions versus variable regions are needed and will provide valuable insights into the mechanisms of SHM and CSR.

WHY ARE REPAIR PATHWAYS MUTAGENIC IN THIS PARTICULAR CASE?

The involvement of *bona fide* repair pathways in the mutagenic pathways SHM and CSR remains puzzling. Under normal circumstances, MMR and BER are high fidelity processes that are important guardians of genomic integrity. Why do they develop a mutagenic function in CSR and SHM, and how can this function be restricted to B-cells?

Germinal center B-cells undergoing SHM and CSR are mostly in the G1 phase of the cell cycle [152]. Canonical MMR that would lead to faithful repair of mismatches can therefore be excluded. The mutagenic properties of non-canonical MMR outside of S-phase have been studied [117, 137, 142]. Its occurrence in non-B-cells seems to be undesirable under normal circumstances. Thanks to alternative mechanism like BER that addresses some mismatches outside of S phase, non-canonical MMR is likely to play a minor role in cells. In B-cells, however, the situation is different, most likely because there is a high accumulation of lesions that can be addressed by MMR.

The processing of uracils by BER is a highly-efficient process. There are four glycosylases in mammals UNG, SMUG1, MDB4 and thymine DNA glycosylase TDG [35], but only UNG has been linked to SHM and CSR. As mentioned in the introduction, UNG has an extremely high turnover rate and is therefore very efficient in removing uracils from DNA. The same explanation for the mutagenic action of MMR in B-cells can apply here. AID creates an unusual accumulation of uracils in close proximity. This could possibly lead to overwhelming of the otherwise very efficient BER.

The case of SHM and CSR constitutes a unique situation, because it involves an accumulation of lesions that can be addressed by two repair pathways simultaneously. The two pathways are normally prevented from encountering each other, as MMR is tightly-coupled to replication, whereas the BER enzyme, TDG for example, is specifically degraded during S-phase.

AID is the only B-cell specific factor known that is involved in these mutagenic processes. It is therefore conceivable that the B-cell specific mutagenicity of the repair pathways is found at that step of regulation. Nevertheless, it is still somewhat unanticipated that the lesions are not repaired faithfully by the two repair pathways involved.

Several hypotheses are under discussion to explain the inefficient repair of the uracils and the induction of dangerous and normally-undesirable DSBs. One or more factors of the repair pathways might be absent or present to a lesser extent in B-cells. One candidate is polymerase β, which has indeed been shown to be down-regulated in activated B-cells [49]. B-cells completely lacking polymerase β show a slight increase in CSR to some isotypes [520].

A possible explanation for the tolerance of B-cells towards DSBs came from the Dalla-Favera group that found reduced p53 expression in activated B-cells [521]. Limited signaling of the important tumor suppressor p53 in an environment of high mutagenesis and frequent DNA rearrangement seems very dangerous though. Accordingly, the incidence of lymphomas is particularly high in B-cells.

B-CELL LYMPHOMAS

Lymphomas are a group of tumors originating from lymphocytes. 95% of all lymphomas are of B-cell origin [522]. The biggest subgroup of Non-Hodgkin lymphomas are the mature B-cell lymphomas. They comprise among others the groups of diffuse large B-cell lymphoma and Burkitt's lymphoma (BL). Most of these cancers are initiated by translocations of proto-oncogenes. Depending on the location of the translocation and on the type of proto-oncogene, its carcinogenic effect can be due to the removal of negative regulators or to the addition of a strong enhancer or active promoter [523].

A hallmark of mature B cell lymphomas is the involvement of at least one *Ig* gene in the transformation-initiating translocation [524-528]. BL for example is characterized by the translocation of *c-Myc* to an immunoglobulin gene locus, in most cases *IgH* [523]. This leads to the erroneous regulation of *c-Myc* and promotes tumor formation through its effect on proliferation [528]. Other examples are the translocation of the *Ig* locus to *Bcl-1*, a controller of the cell cycle or to *Bcl-2*, an inhibitor of apoptosis [529].

The frequent involvement of the *Ig* locus suggests an involvement of the *Ig*-modifying mechanism V(D)J recombination and the mature B-cell specific processes SHM and CSR. Indeed, it has been shown that deregulated expression of the SHM- and CSR-initiating enzyme AID, together with a *p53* mutation, causes extensive genome instability, resulting in the induction of B cell lymphomas [224]. AID has also been shown to act on off-target sequences that share AID-target sequence similarity throughout the genome. It has been shown for example that AID frequently attacks the *c-Myc* sequence when it is transcribed [530], explaining the involvement of *c-Myc* in many translocation events.

Apart from AID, also downstream effectors of SMH and CSR can lead to tumorigenesis when deregulated. UNG-deficient mice show only a slight increase in mutation frequency, but they are susceptible to developing B-cell lymphomas [62]. Deficiencies in NHEJ repair enzymes have also been linked to malignant transformations involving translocations [176, 531, 532]

In line with this, we observed translocations to genomic DNA and to the GFP control plasmid in our *in vivo* system. The incidence of translocations was higher in MMR-depleted cells. This indicates that MMR normally provides some protection against non-compatible ligations, probably by resolving imperfect microhomology-mediated joining of the two ends.

A better understanding of the mechanism leading to B-cell specific DNA rearrangements and mutagenesis is essential for the comprehension of tumorigenesis in B-cells and will open new ways to treat these tumors.

THE ROLE OF DNA METHYLATION IN THE REGULATION OF *VTG* EXPRESSION

We could show that vitellogenin expression is induced upon β-estradiol treatment in LMH/2A cells, but that this is happening independently of detectable changes in DNA methylation of the four CpGs that were reported earlier to be demethylated upon hormone treatment [332]. The fact that we see an accumulation of the TET oxidation product 5hmC at these CpGs confirms that dynamic changes at the level of chromatin are taking place. Whether the appearance of 5hmC is a consequence of increased accessibility for different DNA binding proteins or actually causal for the regulation of expression remains to be determined. TET2 overexpression experiments point towards a role for the dioxygenase in the transcriptional regulation of *VTG*, but they need confirmation from knockdown experiments. We have not yet succeeded in

efficiently downregulating the TETs at the protein level using specific siRNAs, despite the fact that mRNA levels were significantly reduced.

The observation that we could induce *VTG* transcription without affecting DNA methylation puzzled us at first. One limitation to keep in mind is the involvement of a PCR step that can lead to distortion of the sequencing results. Moreover, there is the possibility that the *VTG* transcripts originated from only a few cells that underwent demethylation; the unmethylated sites of these genomes would have escaped detection in sequencing of the entire cellular pool. Alternatively, the signal could have arisen from the cells that acquired 5hmC. The question remains why we do not see at least the same low levels of demethylation as we see 5hmC-signal. If 5hmC was just an intermediate step in demethylation, we should see oxidation events that go further than hydroxymethylation, displaying as unmodified cytosines in bisulfite-conversion assays. In contrast, the oxidation events taking place at the *VTG* enhancer seem to stop specifically at 5hmC, pointing towards a regulatory role of the base, beyond its function as an intermediate in DNA demethylation.

Since this is the first time that methylation dynamics of the *VTG* enhancer were analyzed by bisulfite sequencing, we cannot rule out the possibility that another (yet uncharacterized) modification on cytosine is present, mimicking unmethylated cytosine in *Hpa*II digests and chemical sequencing, but 5mC in bisulfite conversion.

The possibility that DNA demethylation is a mere consequence of the activation of transcription through other means is still conceivable. In the early studies, this was considered very likely, due to the fact that demethylation, as assessed by *Hpa*II digests, was detectable only after 24 h of E_2 treatment, whereas transcription started as soon as 4 h after hormone administration. These studies, however, ignored the fact that hemi-methylated DNA was, similarly to fully- methylated, refractory to cleavage by *Hpa*II. But, hemi-methylation of DNA could display distinct characteristics when it comes to transcription factor binding or chromatin remodeling. The method established by Maxam and Gilbert, and Church and Gilbert to chemically sequence DNA [507, 508], re-opened the discussion about the temporal sequence of transcription initiation and DNA demethylation. An observation made when different tissues were analyzed for their inducibility and methylation dynamics, strongly suggests a causal function of DNA demethylation. Erythrocytes did not express *VTG* when treated with E_2 and there was no demethylation of the enhancer. This was not surprising, since erythrocytes also do not express ERs. Cells of the oviduct, however, do express ERα, but *VTG* is not inducible. Still, the demethylation event was observed in these cells. This indicates that general estrogen-responsiveness is the key prerequisite for the demethylation and not transcription *per se* [332, 503]. This phenomenon also implies that there must be another layer of control that prevents transcription in oviduct cells, despite the presence of ERα and E_2. Whether this layer of control involves methylated CpGs at other sites or other mechanisms of control is not known to date.

CpG **b** shows an interesting behavior in rooster liver, in that it is present in a hemi-methylated state prior to hormone treatment. The presence of hemi-methylated sites in DNA is difficult to explain; first, they are targets for DNMT1, which should convert them to fully-methylated CpGs. Second, replication of hemi-methylated DNA would result in the loss of methylation in 50% of the progeny molecules. It is conceivable though that hemi-methylated sites could persist in DNA of non-replicating cells when protected by e.g. specific binding proteins. In LMH/2A cells, both strands are 50% methylated, most likely representing a CpG that is 50% unmethylated and 50% fully-methylated in a population of cells. The factor or circumstance that protects the hemi-methylated state in rooster liver thus seems to be absent in the cell line. The methylation state at CpG **b** does not seem to confer any advantage to the cell, causing one cell population to outgrow the other.

The binding of ERα to its response element is insensitive to methylation of the two CpG sites present in the *VTG* ERE. Another factor (X) that reconises the ERE, however, binds with lower affinity if the site is methylated. The reporter assay revealed no significant influence of ERE-methylation though, arguing against an important role of factor X binding at the ERE. Interestingly, we detected a shift with a similar mobility on other CpG sites in the *VTG* enhancer/promoter region (CpG **b**, **a**, and 7). Moreover, we observed that methylation-responsiveness of factor X was different on the different CpG substrates, particularly on CpG7. We could compete the binding of the factor at the ERE with an excess of CpG7 or of ΔG that also displays high affinity for factor X, but almost no binding of ERα. These findings prompted us to conclude that it is indeed the same factor(s) that binds different sites with different affinities and binding behavior. The sequences surrounding the different CpGs show no obvious similarity, but CpG7 is located in an E-box and the deletion of the G in the spacer region of the ERE gives rise to in an E-box consensus sequence too.

HELIX-LOOP-HELIX TRANSCRIPTION FACTORS

The binding of the same factor to several CpG-containing sequences of the *VTG* enhancer/promoter attracted our interest, because the factor displayed the strongest affinity for CpG7 and ΔG substrates, both of which contain an E-box consensus sequence. E-box binding proteins form a large superfamily of basic helix-loop-helix (bHLH) factors. bHLH proteins are small, dimeric transcription factors reported to be important regulators of embryonic development. They consist of an N-terminal basic α-helix, responsible for the binding to the E-box motif. The C-terminus forms the HLH structure that facilitates interactions with other members of the family [533]. bHLH proteins are known for their versatility, due to many different combinations of homo- and heterodimers that can be formed depending on the developmental stage and the cell type. This system enables the cell to react flexibly to stimuli with a limited number of proteins. The different combinations of bHLH factors were moreover shown to bind distinct E-box consensus sequences with different affinities [534].

The bHLH proteins are subdivided into many families, depending on the additional domains they possess. An interesting group of bHLH factors building a network around the transcription factor MAX contain an additional leucine zipper domain that assists in dimerization and DNA binding. Leucine zippers contain an amphiphilic α-helical structure consisting of six heptad repeats with leucine residues in the hydrophobic core of the helix. Members of this group of bHLH factors include the different MYCs, MAD and MLX. This family received a great deal of attention, because the *MYC* genes are very frequently mutated in cancer [359], as discussed above on the example of B-cell lymphomas. Depending on their dimerization partners, MYC transcription factors can promote transcription or inhibit it. Methylation-sensitive DNA binding of c-MYC has been reported in combination with certain bHLH proteins, but plays no role in other combinations [535, 536].

When we examined the proteins binding the ERE in our nuclear extracts on denaturing SDS gels, we detected several bands, corresponding to proteins of molecular size of 20-50 kDa. This is consistent with the size of most HLH proteins and considerably smaller than the NHPs published to bind to the ERE of chicken *VTG* [509]. Even more evidence for the involvement of HLH proteins came from the mass spectrometric analysis of the proteins binding the wt and the ΔG ERE in LMH/2A- , as well as in HeLa NEs. We could identify several helix-loop-helix transcription factors from both cell lines; MLX, MLXIP and MAX, in HeLa NEs additionally MLXIPL. The characterization of the role of these factors in transcriptional control with respect to their methylation-sensitivity will be very interesting and might lead to an improved understanding of the role of the MAX network in development and possibly also in malignant transformation.

TEN-ELEVEN TRANSLOCATION PROTEINS AND DNA DEMETHYLATION

Depletion or knockout of TETs in genome-wide studies provided plausible evidence for an important role of the dioxygenases in DNA demethylation during development. There are, however, still a number of open questions concerning their role in somatic cells, their precise targeting and - last but not least - their exact mechanism of action. Considering that 5mC occurs almost exclusively in the symmetrical CpG context, a global wave of demethylation, involving BER, seems to be a very risky task. Each methylated CpG site could potentially be converted into a DSB. We showed in our study (Results I, [517]) the danger of an accumulation of BER intermediates. During CSR, closely-spaced uracils processed by BER and MMR can give rise to DSBs. The situation in the CpG context is even more acute, as the lesions are directly opposite each other if processed simultaneously. The involvement of several oxidation steps is a genuine way of temporally spacing the base excision events. It has, however, not been shown yet, whether the TET enzymes act symmetrically on CpGs and whether they carry out the three oxidation steps in a processive manner, from 5mC to 5caC and excision, without leaving the DNA. Crystal structures of NgTET1 and human TET2 bound to DNA suggest a base flipping mechanism reminiscent of glycosylases, which hints towards asymmetrical oxidations. The data

present evidence for higher turnover when substrates are oxidized processively from 5mC to 5caC [432-434, 437]. However, the relatively high levels of 5hmC in cells, compared to 5fC and 5caC, argue against this. Another study that set out to investigate if TET-mediated oxidation steps occur symmetrically made use of the chemical reactivity of the formyl group of 5fC to selectively attach a fluorescent pyrene group. The authors observed predominantly pyrene modification of both strands, prompting them to conclude not only that the oxidation is symmetrical, but also that the oxidation steps are non-processive, due to the detectable presence of 5fC [537]. These studies were followed by investigations of the mechanism of action of the TETs using high performance liquid chromatography (HPLC) and mass spectrometry (MS). While one group came to the conclusion that TET2 carries out iterative oxidations to 5caC without releasing the DNA in between [538], another group objected, postulating that TETs act in a distributive manner [539]. If symmetrical 5fC or 5caC generation does indeed occur in DNA, it is not certain that their processing would lead to the formation of a DSB, as there is an additional level of control at the BER step. The analysis of symmetrical 5caC substrates with a recombinant TDG-BER machinery resulted in *Hpa*II cleavable sites, but only marginal levels of DSB, indicating an efficient processing of the sites without critically damaging the DNA. The group concluded that TDG interacts tightly with AP-sites, thereby inhibiting simultaneous processing of both strands. This finding indicates that there is an additional level of protection against DSBs during TET-mediated demethylation, which is provided by BER [95]. Despite the great research interest in the field of active DNA demethylation, there are still a number of open questions and contradictory findings. The analysis of the mechanism of demethylation *in vivo* is difficult, because the methods used to date, HPLC, MS [540], antibody-based [471] or using the chemical reactivity of formyl groups [537, 541, 542], cannot distinguish single sites and cannot draw conclusions about both strands at one site. They yield averages, making it difficult to assess the exact mode of action of the enzymes involved. More studies are needed that would differentiate between the two strands at single sites, using for example sequencing methods on hairpin substrates, auch as suggested in 'additional observations'.

Also, the specific targeting of the demethylation machinery requires further investigation. It is important during the global waves of demethylation, for example to protect imprinted genes, but even more so in somatic cells, when single CpGs are addressed. Another layer of complexity concerns the finding that the oxidation machinery fails to proceed beyond the generation of 5hmC in some cases, for example in the brain, where levels of 5hmC are exceptionally high, while in other cases a complete demethylation might be the desired outcome.

TEN-ELEVEN TRANSLOCATION PROTEINS IN CANCER

That TET proteins play a role in cancer development has been clear from the start, as they were discovered by being involved in a translocation event in myeloid cancer [418, 446]. Since then, a substantial amount of evidence has been gathered that strengthens the role of TETs (mostly

TET2) in hematopoietic malignancies. *TET2* has been found to be the most frequently-mutated gene in myelodysplastic syndromes (MDS), with up to 30% penetrance. *TET2* mutations are found in the majority of cells in the bone marrow of patients, indicating that these mutations occur early in disease development [447]. However, no reliable prognostic value could be derived from high-throughput sequencing of patients' DNA [448]. Also, the mechanism by which TET2 acts as a tumor suppressor remains enigmatic.

Several groups studied knockout mouse models to assess the role of *TETs* in malignant transformation. It was found that loss of *TET2* in hematopoietic cells lead to extensive proliferation, resulting in an enlargement of the hematopoietic stem cell compartment due to increased self-renewal. Also, heterozygous HSCs showed increased proliferation, suggesting that *TET2* might be haploinsufficient [450, 543]. A striking reduction of 5hmC levels and concomitant increase of 5mC were observed in HSCs of TET2-deficient mice and differentiation balances were altered towards monocytic/granulocytic lineages [449]. The importance of *TETs* as tumour suppressors was further strengthened by the finding that mutations in isocitrate dehydrogenases (*IDH1/2*), enzymes that convert isocitrate to the TET substrate 2-oxoglutarate, were found to cause similar cancers as *TET* mutations, which were characterized by global DNA hypermethylation. *IDH1/2* mutations were additionally shown to be mutually exclusive with *TET2* mutations in cancer [544].

The observation that *TET2* mutations are more often found in cancers compared to defects in *TET1* or *TET3*, indicates different roles and distribution of the *TETs*. Despite of the considerable amount of clinical data on the involvement of the TETs in cancer, the exact mechanism and the key genes that are deregulated in patients remain to be characterized.

CONCLUDING REMARKS & FUTURE PERSPECTIVES

The involvement of several repair pathways in antibody diversification and regulation of gene expression and chromatin dynamics shows their great importance, much beyond repair of damage arising through endogenous or exogenous agents. Instead of only reacting to damage, they actively modulate cellular processes, thereby enabling higher forms of life. The term 'repair pathways' clearly undervalues these versatile and highly complex mechanisms; a better term might be DNA modulating pathways.

The usage of DNA substrates containing nicks and uracils to investigate the process of CSR enabled us to gain insights into the exact way of processing of the lesions. By dispensing with AID, we reduced the complexity of the events by knowing exactly where the lesions we started with were located. The puzzling results obtained with the more physiological substrates containing uracils on both strands merit further attention, as they imply that simple nicks may not be identical to intermediates arising during processing of damaged bases. Thus, substrates containg uracils rather than nicks must be reexamined to reveal the extent of accessibility of repair intermediates to other pathways, such as MMR. Experiments with substrates containing uracils on both strands would moreover provide valuable insights into the mechanism of CSR in B-cells, as it mimics the situation present in S regions acted-upon by AID.

Another issue worth addressing in the future is the reason for the unusually inefficient BER during antibody diversification. The above-mentioned saturation of the repair pathway by the creation of a high density of targets in close proximity is a conceivable explanation, it has, however, never been experimentally addressed. Downregulation or modification of key enzymes of the pathway in germinal center B-cells is also under discussion.

There is still a great deal to learn about CSR and the enzymes and pathways involved. It is a very unique and fascinating process and a valuable model system to investigate the interplay of several DNA metabolic pathways. The finding that CSR is frequently involved in the formation of malignancies makes its detailed understanding particularly relevant.

In addition to *bona fide* DNA repair and antibody diversification, BER appears to have an important role in DNA demethylation and the control of gene expression. Epigenetic processes are extremely challenging to decipher in cells, as they are complex and involve many different players and consecutive modifications of chromatin. The order of events leading to the activation or shut-down of a specific gene is difficult to address and might vary from gene to gene. The fact that methylation inhibits TF binding in some cases, either directly or *via* the binding of MBD proteins, argues for an important causal role of DNA methylation in the control of gene expression. There are, however, also numerous examples, where DNA demethylation is reported to occur after the onset of transcription. In hormone-regulated genes, as discussed in this thesis, the binding of the activated TF is very likely to be the first event, serving to target the DNA modifying machinery to the specific site. The *VTG* gene represents a valuable model system

to study the events on chromatin, leading to expression of the gene, as it is well characterized and inducible by estradiol treatment.

The data presented in this thesis illustrate the diverse functions of eukaryotic repair pathways. They show the great potential of these versatile mechanisms, but also point out the danger they pose to cells and multicellular organisms. The flexibility of DNA modifications and rearrangements, enabling an efficient immune protection and differential gene expression in higher eukaryotes, goes at the expense of an increased risk for DNA damage and uncontrolled proliferation, potentially leading to tumor formation. A deep understanding of the mechanisms involved and their regulation and targeting is of outmost importance for successful cancer prevention and treatment, without disturbing the beneficial aspects of flexible DNA metabolism.

REFERENCES

1. Hahn, W.C. and R.A. Weinberg, *Modelling the molecular circuitry of cancer.* Nat Rev Cancer, 2002. **2**(5): p. 331-41.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
4. Soulas-Sprauel, P., et al., *V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining.* Oncogene, 2007. **26**(56): p. 7780-91.
5. Dudley, D.D., et al., *Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences.* Advances in immunology, 2005. **86**: p. 43-112.
6. Marmur, J. and P. Doty, *Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature.* J Mol Biol, 1962. **5**: p. 109-18.
7. Wartell, R.M. and A.S. Benight, *Thermal-Denaturation of DNA-Molecules - a Comparison of Theory with Experiment.* Physics Reports-Review Section of Physics Letters, 1985. **126**(2): p. 67-107.
8. Yakovchuk, P., E. Protozanova, and M.D. Frank-Kamenetskii, *Base-stacking and base-pairing contributions into thermal stability of the DNA double helix.* Nucleic Acids Res, 2006. **34**(2): p. 564-74.
9. Wei, Y.H., *Oxidative stress and mitochondrial DNA mutations in human aging.* Proc Soc Exp Biol Med, 1998. **217**(1): p. 53-63.
10. Monsalve, M., et al., *Mitochondrial dysfunction in human pathologies.* Front Biosci, 2007. **12**: p. 1131-53.
11. Chance, B., H. Sies, and A. Boveris, *Hydroperoxide metabolism in mammalian organs.* Physiol Rev, 1979. **59**(3): p. 527-605.
12. Dizdaroglu, M., et al., *Free radical-induced damage to DNA: mechanisms and measurement.* Free Radic Biol Med, 2002. **32**(11): p. 1102-15.
13. Valko, M., et al., *Free radicals, metals and antioxidants in oxidative stress-induced cancer.* Chem Biol Interact, 2006. **160**(1): p. 1-40.
14. van Loon, B., E. Markkanen, and U. Hubscher, *Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine.* DNA Repair (Amst), 2010. **9**(6): p. 604-16.
15. Valko, M., et al., *Role of oxygen radicals in DNA damage and cancer incidence.* Mol Cell Biochem, 2004. **266**(1-2): p. 37-56.
16. Greenman, C., et al., *Patterns of somatic mutation in human cancer genomes.* Nature, 2007. **446**(7132): p. 153-8.
17. Frederico, L.A., T.A. Kunkel, and B.R. Shaw, *A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy.* Biochemistry, 1990. **29**(10): p. 2532-7.
18. Shapiro, R. and M. Danzig, *Acidic hydrolysis of deoxycytidine and deoxyuridine derivatives. The general mechanism of deoxyribonucleoside hydrolysis.* Biochemistry, 1972. **11**(1): p. 23-9.
19. Shen, J.C., W.M. Rideout, 3rd, and P.A. Jones, *The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA.* Nucleic Acids Res, 1994. **22**(6): p. 972-6.
20. Duncan, B.K. and J.H. Miller, *Mutagenic deamination of cytosine residues in DNA.* Nature, 1980. **287**(5782): p. 560-1.
21. Ehrlich, M., et al., *DNA cytosine methylation and heat-induced deamination.* Biosci Rep, 1986. **6**(4): p. 387-93.
22. Hill-Perkins, M., M.D. Jones, and P. Karran, *Site-specific mutagenesis in vivo by single methylated or deaminated purine bases.* Mutat Res, 1986. **162**(2): p. 153-63.

23. Karran, P. and T. Lindahl, *Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus.* Biochemistry, 1980. **19**(26): p. 6005-11.
24. Lindahl, T. and B. Nyberg, *Rate of depurination of native deoxyribonucleic acid.* Biochemistry, 1972. **11**(19): p. 3610-8.
25. Gates, K.S., *An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals.* Chem Res Toxicol, 2009. **22**(11): p. 1747-60.
26. Doerfler, W., *DNA methylation and gene activity.* Annu Rev Biochem, 1983. **52**: p. 93-124.
27. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.
28. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.
29. Harris, R.S., S.K. Petersen-Mahrt, and M.S. Neuberger, *RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.* Mol Cell, 2002. **10**(5): p. 1247-53.
30. Knisbacher, B.A., D. Gerber, and E.Y. Levanon, *DNA Editing by APOBECs: A Genomic Preserver and Transformer.* Trends Genet, 2016. **32**(1): p. 16-28.
31. Lindahl, T., B. Demple, and P. Robins, *Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase.* EMBO J, 1982. **1**(11): p. 1359-63.
32. Slupphaug, G., et al., *Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase.* Biochemistry, 1995. **34**(1): p. 128-38.
33. Kavli, B., et al., *hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup.* The Journal of biological chemistry, 2002. **277**(42): p. 39926-36.
34. Lindahl, T., *An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues.* Proc Natl Acad Sci U S A, 1974. **71**(9): p. 3649-53.
35. Jacobs, A.L. and P. Schar, *DNA glycosylases: in DNA repair and beyond.* Chromosoma, 2012. **121**(1): p. 1-20.
36. David, S.S., V.L. O'Shea, and S. Kundu, *Base-excision repair of oxidative DNA damage.* Nature, 2007. **447**(7147): p. 941-50.
37. Friedberg, E.C., et al., *DNA repair: from molecular mechanism to human disease.* DNA Repair (Amst), 2006. **5**(8): p. 986-96.
38. Robertson, A.B., et al., *DNA repair in mammalian cells: Base excision repair: the long and short of it.* Cellular and molecular life sciences : CMLS, 2009. **66**(6): p. 981-93.
39. Cortazar, D., et al., *Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability.* Nature, 2011. **470**(7334): p. 419-23.
40. Kubota, Y., et al., *Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein.* The EMBO journal, 1996. **15**(23): p. 6662-70.
41. Frosina, G., et al., *Two pathways for base excision repair in mammalian cells.* The Journal of biological chemistry, 1996. **271**(16): p. 9573-8.
42. Klungland, A. and T. Lindahl, *Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1).* The EMBO journal, 1997. **16**(11): p. 3341-8.
43. Kim, K., S. Biade, and Y. Matsumoto, *Involvement of flap endonuclease 1 in base excision DNA repair.* The Journal of biological chemistry, 1998. **273**(15): p. 8842-8.
44. Matsumoto, Y., K. Kim, and D.F. Bogenhagen, *Proliferating cell nuclear antigen-dependent abasic site repair in Xenopus laevis oocytes: an alternative pathway of base excision DNA repair.* Molecular and cellular biology, 1994. **14**(9): p. 6187-97.
45. Podlutsky, A.J., et al., *Human DNA polymerase beta initiates DNA synthesis during long-patch repair of reduced AP sites in DNA.* EMBO J, 2001. **20**(6): p. 1477-82.

46. Pascucci, B., et al., *Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon.* J Biol Chem, 1999. **274**(47): p. 33696-702.
47. Sleeth, K.M., R.L. Robson, and G.L. Dianov, *Exchangeability of mammalian DNA ligases between base excision repair pathways.* Biochemistry, 2004. **43**(40): p. 12924-30.
48. Haushalter, K.A., et al., *Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors.* Curr Biol, 1999. **9**(4): p. 174-85.
49. Stavnezer, J., J.E. Guikema, and C.E. Schrader, *Mechanism and regulation of class switch recombination.* Annual review of immunology, 2008. **26**: p. 261-92.
50. Krokan, H.E., et al., *Base excision repair of DNA in mammalian cells.* FEBS Lett, 2000. **476**(1-2): p. 73-7.
51. Kavli, B., et al., *Uracil in DNA--general mutagen, but normal intermediate in acquired immunity.* DNA repair, 2007. **6**(4): p. 505-16.
52. Pettersen, H.S., et al., *UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation.* Nucleic Acids Res, 2011. **39**(19): p. 8430-44.
53. Fischer, F., K. Baerenfaller, and J. Jiricny, *5-Fluorouracil is efficiently removed from DNA by the base excision and mismatch repair systems.* Gastroenterology, 2007. **133**(6): p. 1858-68.
54. Mol, C.D., et al., *Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis.* Cell, 1995. **80**(6): p. 869-78.
55. Kavli, B., et al., *Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase.* EMBO J, 1996. **15**(13): p. 3442-7.
56. Nilsen, H., et al., *Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene.* Nucleic Acids Res, 1997. **25**(4): p. 750-5.
57. Haug, T., et al., *Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase.* Nucleic Acids Res, 1998. **26**(6): p. 1449-57.
58. Imai, K., et al., *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nature immunology, 2003. **4**(10): p. 1023-8.
59. Otterlei, M., et al., *Post-replicative base excision repair in replication foci.* EMBO J, 1999. **18**(13): p. 3834-44.
60. Akbari, M., et al., *Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells.* Nucleic acids research, 2004. **32**(18): p. 5486-98.
61. Nilsen, H., et al., *Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication.* Molecular cell, 2000. **5**(6): p. 1059-65.
62. Nilsen, H., et al., *Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas.* Oncogene, 2003. **22**(35): p. 5381-6.
63. Obeid, S., et al., *Replication through an abasic DNA lesion: structural basis for adenine selectivity.* EMBO J, 2010. **29**(10): p. 1738-47.
64. Rada, C., et al., *Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice.* Current biology : CB, 2002. **12**(20): p. 1748-55.
65. Durandy, A., *Hyper-IgM syndromes: a model for studying the regulation of class switch recombination and somatic hypermutation generation.* Biochem Soc Trans, 2002. **30**(4): p. 815-8.
66. Revy, P., et al., *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2).* Cell, 2000. **102**(5): p. 565-75.
67. Korthauer, U., et al., *Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM.* Nature, 1993. **361**(6412): p. 539-41.
68. Hare, J.T. and J.H. Taylor, *One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair.* Proc Natl Acad Sci U S A, 1985. **82**(21): p. 7350-4.
69. Modrich, P., *Mismatch correction.* Basic Life Sci, 1986. **38**: p. 303-10.

70. Radman, M. and R. Wagner, *Mismatch repair in Escherichia coli.* Annu Rev Genet, 1986. **20**: p. 523-38.
71. Wang, R.Y., et al., *Heat- and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA.* Biochim Biophys Acta, 1982. **697**(3): p. 371-7.
72. Brown, T.C. and J. Jiricny, *A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine.* Cell, 1987. **50**(6): p. 945-50.
73. Wiebauer, K. and J. Jiricny, *In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells.* Nature, 1989. **339**(6221): p. 234-6.
74. Wiebauer, K. and J. Jiricny, *Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells.* Proc Natl Acad Sci U S A, 1990. **87**(15): p. 5842-5.
75. Neddermann, P. and J. Jiricny, *The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells.* J Biol Chem, 1993. **268**(28): p. 21218-24.
76. Neddermann, P., et al., *Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase.* J Biol Chem, 1996. **271**(22): p. 12767-74.
77. Neddermann, P. and J. Jiricny, *Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1642-6.
78. Wang, Z. and D.W. Mosbaugh, *Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase.* J Biol Chem, 1989. **264**(2): p. 1163-71.
79. Bennett, M.T., et al., *Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability.* J Am Chem Soc, 2006. **128**(38): p. 12510-9.
80. Hardeland, U., et al., *The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs.* Nucleic Acids Res, 2003. **31**(9): p. 2261-71.
81. Liu, P., A. Burdzy, and L.C. Sowers, *Repair of the mutagenic DNA oxidation product, 5-formyluracil.* DNA Repair (Amst), 2003. **2**(2): p. 199-210.
82. Maiti, A., et al., *Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition.* Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8890-5.
83. Barrett, T.E., et al., *Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions.* Cell, 1998. **92**(1): p. 117-29.
84. Barrett, T.E., et al., *Crystal structure of a thwarted mismatch glycosylase DNA repair complex.* EMBO J, 1999. **18**(23): p. 6599-609.
85. Hardeland, U., et al., *Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis.* J Biol Chem, 2000. **275**(43): p. 33449-56.
86. Hardeland, U., et al., *Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover.* EMBO J, 2002. **21**(6): p. 1456-64.
87. Steinacher, R. and P. Schar, *Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation.* Curr Biol, 2005. **15**(7): p. 616-23.
88. Hardeland, U., et al., *Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2.* Nucleic Acids Res, 2007. **35**(11): p. 3859-67.
89. Slenn, T.J., et al., *Thymine DNA glycosylase is a CRL4Cdt2 substrate.* J Biol Chem, 2014. **289**(33): p. 23043-55.
90. Zhong, W., et al., *CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing.* Nature, 2003. **423**(6942): p. 885-9.
91. Havens, C.G. and J.C. Walter, *Mechanism of CRL4(Cdt2), a PCNA-dependent E3 ubiquitin ligase.* Genes Dev, 2011. **25**(15): p. 1568-82.
92. Li, Z., et al., *Gadd45a promotes DNA demethylation through TDG.* Nucleic Acids Res, 2015. **43**(8): p. 3986-97.

93. Zhu, B., et al., *5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5135-9.
94. Pidugu, L.S., et al., *Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase.* Biochemistry, 2016. **55**(45): p. 6205-6208.
95. Weber, A.R., et al., *Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism.* Nat Commun, 2016. **7**: p. 10806.
96. Zhang, L., et al., *Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA.* Nat Chem Biol, 2012. **8**(4): p. 328-30.
97. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.* Science, 2011. **333**(6047): p. 1303-7.
98. Cortellino, S., et al., *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair.* Cell, 2011. **146**(1): p. 67-79.
99. Hendrich, B. and A. Bird, *Identification and characterization of a family of mammalian methyl-CpG binding proteins.* Mol Cell Biol, 1998. **18**(11): p. 6538-47.
100. Zhang, Y., et al., *Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation.* Genes Dev, 1999. **13**(15): p. 1924-35.
101. Saito, M. and F. Ishikawa, *The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2.* J Biol Chem, 2002. **277**(38): p. 35434-9.
102. Hendrich, B., et al., *The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites.* Nature, 1999. **401**(6750): p. 301-4.
103. Zhu, B., et al., *5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence.* Nucleic Acids Res, 2000. **28**(21): p. 4157-65.
104. Rai, K., et al., *DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45.* Cell, 2008. **135**(7): p. 1201-12.
105. Millar, C.B., et al., *Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice.* Science, 2002. **297**(5580): p. 403-5.
106. Arana, M.E. and T.A. Kunkel, *Mutator phenotypes due to DNA replication infidelity.* Seminars in cancer biology, 2010. **20**(5): p. 304-11.
107. Peltomaki, P., *Role of DNA mismatch repair defects in the pathogenesis of human cancer.* J Clin Oncol, 2003. **21**(6): p. 1174-9.
108. Wimmer, K. and J. Etzler, *Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg?* Human genetics, 2008. **124**(2): p. 105-22.
109. Liu, B., et al., *Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients.* Nat Med, 1996. **2**(2): p. 169-74.
110. Colussi, C., et al., *The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool.* Current biology : CB, 2002. **12**(11): p. 912-8.
111. Bignami, M., I. Casorelli, and P. Karran, *Mismatch repair and response to DNA-damaging antitumour therapies.* European journal of cancer, 2003. **39**(15): p. 2142-9.
112. Harfe, B.D. and S. Jinks-Robertson, *DNA mismatch repair and genetic instability.* Annu Rev Genet, 2000. **34**: p. 359-399.
113. Paques, F. and J.E. Haber, *Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 349-404.
114. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annual review of biochemistry, 2005. **74**: p. 681-710.
115. Bardwell, P.D., et al., *Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice.* Nature immunology, 2004. **5**(2): p. 224-9.
116. Chahwan, R., et al., *AIDing antibody diversity by error-prone mismatch repair.* Seminars in immunology, 2012. **24**(4): p. 293-300.

117. Pena-Diaz, J., et al., *Noncanonical mismatch repair as a source of genomic instability in human cells.* Molecular cell, 2012. **47**(5): p. 669-80.
118. Schanz, S., et al., *Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(14): p. 5593-8.
119. Alvino, E., et al., *High-frequency microsatellite instability is associated with defective DNA mismatch repair in human melanoma.* J Invest Dermatol, 2002. **118**(1): p. 79-86.
120. Kunkel, T.A., *Nucleotide repeats. Slippery DNA and diseases.* Nature, 1993. **365**(6443): p. 207-8.
121. Wagner, R., Jr. and M. Meselson, *Repair tracts in mismatched DNA heteroduplexes.* Proceedings of the National Academy of Sciences of the United States of America, 1976. **73**(11): p. 4135-9.
122. Jiricny, J., et al., *A human 200-kDa protein binds selectively to DNA fragments containing G.T mismatches.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(23): p. 8860-4.
123. Hughes, M.J. and J. Jiricny, *The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities.* The Journal of biological chemistry, 1992. **267**(33): p. 23876-82.
124. Dzantiev, L., et al., *A defined human system that supports bidirectional mismatch-provoked excision.* Molecular cell, 2004. **15**(1): p. 31-41.
125. Cho, W.K., et al., *ATP alters the diffusion mechanics of MutS on mismatched DNA.* Structure, 2012. **20**(7): p. 1264-74.
126. Claverys, J.P. and S.A. Lacks, *Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria.* Microbiological reviews, 1986. **50**(2): p. 133-65.
127. Holmes, J., Jr., S. Clark, and P. Modrich, *Strand-specific mismatch correction in nuclear extracts of human and Drosophila melanogaster cell lines.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(15): p. 5837-41.
128. Thomas, D.C., J.D. Roberts, and T.A. Kunkel, *Heteroduplex repair in extracts of human HeLa cells.* The Journal of biological chemistry, 1991. **266**(6): p. 3744-51.
129. Ghodgaonkar, M.M., et al., *Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair.* Molecular cell, 2013. **50**(3): p. 323-32.
130. Repmann, S., M. Olivera-Harris, and J. Jiricny, *Influence of oxidized purine processing on strand directionality of mismatch repair.* J Biol Chem, 2015. **290**(16): p. 9986-99.
131. Karran, P., *Mechanisms of tolerance to DNA damaging therapeutic drugs.* Carcinogenesis, 2001. **22**(12): p. 1931-7.
132. Bignami, M., et al., *Unmasking a killer: DNA O(6)-methylguanine and the cytotoxicity of methylating agents.* Mutation research, 2000. **462**(2-3): p. 71-82.
133. York, S.J. and P. Modrich, *Mismatch repair-dependent iterative excision at irreparable O6-methylguanine lesions in human nuclear extracts.* The Journal of biological chemistry, 2006. **281**(32): p. 22674-83.
134. Olivera Harris, M., et al., *Mismatch repair-dependent metabolism of O6-methylguanine-containing DNA in Xenopus laevis egg extracts.* DNA Repair (Amst), 2015. **28**: p. 1-7.
135. Shimada, A., et al., *MutS stimulates the endonuclease activity of MutL in an ATP-hydrolysis-dependent manner.* The FEBS journal, 2013. **280**(14): p. 3467-79.
136. Kadyrov, F.A., et al., *Endonucleolytic function of MutLalpha in human mismatch repair.* Cell, 2006. **126**(2): p. 297-308.
137. Ehrenstein, M.R. and M.S. Neuberger, *Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation.* The EMBO journal, 1999. **18**(12): p. 3484-90.
138. Rada, C., et al., *Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting.* Immunity, 1998. **9**(1): p. 135-41.

139. Schrader, C.E., et al., *Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes.* The Journal of experimental medicine, 1999. **190**(3): p. 323-30.
140. Shen, H.M., et al., *Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice.* Journal of immunology, 2006. **177**(8): p. 5386-92.
141. Vora, K.A., et al., *Severe attenuation of the B cell immune response in Msh2-deficient mice.* J Exp Med, 1999. **189**(3): p. 471-82.
142. Martomo, S.A., W.W. Yang, and P.J. Gearhart, *A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination.* The Journal of experimental medicine, 2004. **200**(1): p. 61-8.
143. Phung, Q.H., et al., *Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein.* The Journal of experimental medicine, 1998. **187**(11): p. 1745-51.
144. Kannouche, P.L., J. Wing, and A.R. Lehmann, *Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage.* Molecular cell, 2004. **14**(4): p. 491-500.
145. Maizels, N., *Immunoglobulin gene diversification.* Annual review of genetics, 2005. **39**: p. 23-46.
146. Muramatsu, M., et al., *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.* Cell, 2000. **102**(5): p. 553-63.
147. Rada, C., J.M. Di Noia, and M.S. Neuberger, *Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation.* Molecular cell, 2004. **16**(2): p. 163-71.
148. Zeng, X., et al., *DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes.* Nature immunology, 2001. **2**(6): p. 537-41.
149. Rogozin, I.B., et al., *Somatic mutation hotspots correlate with DNA polymerase eta error spectrum.* Nature immunology, 2001. **2**(6): p. 530-6.
150. Wiesendanger, M., et al., *Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern.* The Journal of experimental medicine, 2000. **191**(3): p. 579-84.
151. Delbos, F., et al., *Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse.* J Exp Med, 2005. **201**(8): p. 1191-6.
152. Schrader, C.E., et al., *Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair.* Journal of immunology, 2007. **179**(9): p. 6064-71.
153. Caldecott, K.W., *DNA single-strand break repair.* Exp Cell Res, 2014. **329**(1): p. 2-8.
154. Davidovic, L., et al., *Importance of poly(ADP-ribose) glycohydrolase in the control of poly(ADP-ribose) metabolism.* Exp Cell Res, 2001. **268**(1): p. 7-13.
155. Fisher, A.E., et al., *Poly(ADP-ribose) polymerase 1 accelerates single-strand break repair in concert with poly(ADP-ribose) glycohydrolase.* Mol Cell Biol, 2007. **27**(15): p. 5597-605.
156. Kim, Y.J. and D.M. Wilson, 3rd, *Overview of base excision repair biochemistry.* Curr Mol Pharmacol, 2012. **5**(1): p. 3-13.
157. Pommier, Y., et al., *Tyrosyl-DNA-phosphodiesterases (TDP1 and TDP2).* DNA Repair (Amst), 2014. **19**: p. 114-29.
158. Kanaar, R., J.H. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double strand break repair.* Trends in cell biology, 1998. **8**(12): p. 483-9.
159. Lieber, M.R., et al., *Mechanism and regulation of human non-homologous DNA end-joining.* Nature reviews. Molecular cell biology, 2003. **4**(9): p. 712-20.
160. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends Cell Biol, 2016. **26**(1): p. 52-64.
161. Hartlerode, A., et al., *Cell cycle-dependent induction of homologous recombination by a tightly regulated I-SceI fusion protein.* PloS one, 2011. **6**(3): p. e16501.

162. Takata, M., et al., *Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.* EMBO J, 1998. **17**(18): p. 5497-508.
163. Carney, J.P., et al., *The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response.* Cell, 1998. **93**(3): p. 477-86.
164. Stewart, G.S., et al., *The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder.* Cell, 1999. **99**(6): p. 577-87.
165. Lee, J.H. and T.T. Paull, *Activation and regulation of ATM kinase activity in response to DNA double-strand breaks.* Oncogene, 2007. **26**(56): p. 7741-8.
166. Aylon, Y., B. Liefshitz, and M. Kupiec, *The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle.* The EMBO journal, 2004. **23**(24): p. 4868-75.
167. Sugiyama, T., E.M. Zaitseva, and S.C. Kowalczykowski, *A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the Saccharomyces cerevisiae Rad51 protein.* J Biol Chem, 1997. **272**(12): p. 7940-5.
168. Haruta, N., et al., *The Swi5-Sfr1 complex stimulates Rhp51/Rad51- and Dmc1-mediated DNA strand exchange in vitro.* Nat Struct Mol Biol, 2006. **13**(9): p. 823-30.
169. New, J.H., et al., *Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A.* Nature, 1998. **391**(6665): p. 407-10.
170. Chen, P.L., et al., *The BRC repeats in BRCA2 are critical for RAD51 binding and resistance to methyl methanesulfonate treatment.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5287-92.
171. Sharan, S.K., et al., *Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2.* Nature, 1997. **386**(6627): p. 804-10.
172. Jasin, M., *Homologous repair of DNA damage and tumorigenesis: the BRCA connection.* Oncogene, 2002. **21**(58): p. 8981-93.
173. Ogawa, T., et al., *Similarity of the yeast RAD51 filament to the bacterial RecA filament.* Science, 1993. **259**(5103): p. 1896-9.
174. San Filippo, J., P. Sung, and H. Klein, *Mechanism of eukaryotic homologous recombination.* Annual review of biochemistry, 2008. **77**: p. 229-57.
175. Ferguson, D.O. and W.K. Holloman, *Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model.* Proc Natl Acad Sci U S A, 1996. **93**(11): p. 5419-24.
176. Ferguson, D.O., et al., *The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6630-3.
177. Nussenzweig, A., et al., *Requirement for Ku80 in growth and immunoglobulin V(D)J recombination.* Nature, 1996. **382**(6591): p. 551-5.
178. Fischer, A., et al., *Naturally occurring primary deficiencies of the immune system.* Annu Rev Immunol, 1997. **15**: p. 93-124.
179. Stephan, J.L., et al., *Severe combined immunodeficiency: a retrospective single-center study of clinical presentation and outcome in 117 patients.* J Pediatr, 1993. **123**(4): p. 564-72.
180. Moshous, D., et al., *Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency.* Cell, 2001. **105**(2): p. 177-86.
181. Ma, Y., et al., *Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination.* Cell, 2002. **108**(6): p. 781-94.
182. Dynan, W.S. and S. Yoo, *Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids.* Nucleic acids research, 1998. **26**(7): p. 1551-9.
183. Gottlieb, T.M. and S.P. Jackson, *The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen.* Cell, 1993. **72**(1): p. 131-42.

184. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annual review of biochemistry, 2010. **79**: p. 181-211.
185. Weinstock, D.M., E. Brunet, and M. Jasin, *Formation of NHEJ-derived reciprocal chromosomal translocations does not require Ku70.* Nature cell biology, 2007. **9**(8): p. 978-81.
186. Gu, J., et al., *XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps.* The EMBO journal, 2007. **26**(4): p. 1010-23.
187. Boulton, S.J. and S.P. Jackson, *Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance.* Nucleic Acids Res, 1996. **24**(23): p. 4639-48.
188. Liang, F., et al., *Chromosomal double-strand break repair in Ku80-deficient cells.* Proc Natl Acad Sci U S A, 1996. **93**(17): p. 8929-33.
189. Lee, K. and S.E. Lee, *Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining.* Genetics, 2007. **176**(4): p. 2003-14.
190. Ma, J.L., et al., *Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences.* Molecular and cellular biology, 2003. **23**(23): p. 8820-8.
191. Decottignies, A., *Microhomology-mediated end joining in fission yeast is repressed by pku70 and relies on genes involved in homologous recombination.* Genetics, 2007. **176**(3): p. 1403-15.
192. Liang, L., et al., *Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks.* Nucleic acids research, 2008. **36**(10): p. 3297-310.
193. Pan-Hammarstrom, Q., et al., *Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells.* The Journal of experimental medicine, 2005. **201**(2): p. 189-94.
194. Audebert, M., B. Salles, and P. Calsou, *Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining.* The Journal of biological chemistry, 2004. **279**(53): p. 55117-26.
195. Wang, M., et al., *PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways.* Nucleic acids research, 2006. **34**(21): p. 6170-82.
196. Weinfeld, M., et al., *Interaction of DNA-dependent protein kinase and poly(ADP-ribose) polymerase with radiation-induced DNA strand breaks.* Radiation research, 1997. **148**(1): p. 22-8.
197. D'Silva, I., et al., *Relative affinities of poly(ADP-ribose) polymerase and DNA-dependent protein kinase for DNA strand interruptions.* Biochimica et biophysica acta, 1999. **1430**(1): p. 119-26.
198. Gellert, M., *V(D)J recombination: RAG proteins, repair factors, and regulation.* Annu Rev Biochem, 2002. **71**: p. 101-32.
199. Swanson, P.C., *The bounty of RAGs: recombination signal complexes and reaction outcomes.* Immunol Rev, 2004. **200**: p. 90-114.
200. Schatz, D.G. and P.C. Swanson, *V(D)J recombination: mechanisms of initiation.* Annual review of genetics, 2011. **45**: p. 167-202.
201. Peled, J.U., et al., *The biochemistry of somatic hypermutation.* Annual review of immunology, 2008. **26**: p. 481-511.
202. Orthwein, A. and J.M. Di Noia, *Activation induced deaminase: how much and where?* Seminars in immunology, 2012. **24**(4): p. 246-54.
203. Neuberger, M.S. and C. Milstein, *Somatic hypermutation.* Current opinion in immunology, 1995. **7**(2): p. 248-54.
204. Rogozin, I.B., T.A. Kunkel, and Y.I. Pavlov, *Double-strand breaks in DNA during somatic hypermutation of Ig genes: cause or consequence?* Trends in immunology, 2002. **23**(1): p. 12-3.

205. Manis, J.P., M. Tian, and F.W. Alt, *Mechanism and control of class-switch recombination.* Trends in immunology, 2002. **23**(1): p. 31-9.
206. Teng, G. and F.N. Papavasiliou, *Immunoglobulin somatic hypermutation.* Annual review of genetics, 2007. **41**: p. 107-20.
207. Muramatsu, M., et al., *Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells.* The Journal of biological chemistry, 1999. **274**(26): p. 18470-6.
208. Dickerson, S.K., et al., *AID mediates hypermutation by deaminating single stranded DNA.* J Exp Med, 2003. **197**(10): p. 1291-6.
209. Bransteitter, R., et al., *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4102-7.
210. Fukita, Y., H. Jacobs, and K. Rajewsky, *Somatic hypermutation in the heavy chain locus correlates with transcription.* Immunity, 1998. **9**(1): p. 105-14.
211. Bachl, J., et al., *Increased transcription levels induce higher mutation rates in a hypermutating cell line.* J Immunol, 2001. **166**(8): p. 5051-7.
212. Chaudhuri, J., C. Khuong, and F.W. Alt, *Replication protein A interacts with AID to promote deamination of somatic hypermutation targets.* Nature, 2004. **430**(7003): p. 992-8.
213. Shen, H.M. and U. Storb, *Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled.* Proc Natl Acad Sci U S A, 2004. **101**(35): p. 12997-3002.
214. Larijani, M. and A. Martin, *Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID.* Mol Cell Biol, 2007. **27**(23): p. 8038-48.
215. Rogozin, I.B. and N.A. Kolchanov, *Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis.* Biochim Biophys Acta, 1992. **1171**(1): p. 11-8.
216. Conticello, S.G., et al., *DNA deamination in immunity: AID in the context of its APOBEC relatives.* Adv Immunol, 2007. **94**: p. 37-73.
217. Larijani, M., et al., *AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner.* Mol Cell Biol, 2007. **27**(1): p. 20-30.
218. Barreto, V., et al., *C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion.* Mol Cell, 2003. **12**(2): p. 501-8.
219. Ranjit, S., et al., *AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus.* Journal of immunology, 2011. **187**(5): p. 2464-75.
220. Shinkura, R., et al., *Separate domains of AID are required for somatic hypermutation and class-switch recombination.* Nat Immunol, 2004. **5**(7): p. 707-12.
221. Zan, H. and P. Casali, *Regulation of Aicda expression and AID activity.* Autoimmunity, 2013. **46**(2): p. 83-101.
222. Hase, K., et al., *Activation-induced cytidine deaminase deficiency causes organ-specific autoimmune disease.* PloS one, 2008. **3**(8): p. e3033.
223. Dorsett, Y., et al., *MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation.* Immunity, 2008. **28**(5): p. 630-8.
224. Robbiani, D.F., et al., *AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations.* Molecular cell, 2009. **36**(4): p. 631-41.
225. Jiang, C., M.L. Zhao, and M. Diaz, *Activation-induced deaminase heterozygous MRL/lpr mice are delayed in the production of high-affinity pathogenic antibodies and in the development of lupus nephritis.* Immunology, 2009. **126**(1): p. 102-13.
226. Okazaki, I.M., et al., *Constitutive expression of AID leads to tumorigenesis.* The Journal of experimental medicine, 2003. **197**(9): p. 1173-81.
227. de Yebenes, V.G., et al., *miR-181b negatively regulates activation-induced cytidine deaminase in B cells.* The Journal of experimental medicine, 2008. **205**(10): p. 2199-206.

228. Aoufouchi, S., et al., *Proteasomal degradation restricts the nuclear lifespan of AID.* The Journal of experimental medicine, 2008. **205**(6): p. 1357-68.
229. Brar, S.S., M. Watson, and M. Diaz, *Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks.* The Journal of biological chemistry, 2004. **279**(25): p. 26395-401.
230. Ito, S., et al., *Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(7): p. 1975-80.
231. Basu, U., et al., *The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation.* Nature, 2005. **438**(7067): p. 508-11.
232. McBride, K.M., et al., *Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8798-803.
233. McBride, K.M., et al., *Regulation of class switch recombination and somatic mutation by AID phosphorylation.* J Exp Med, 2008. **205**(11): p. 2585-94.
234. Pasqualucci, L., et al., *PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells.* Proc Natl Acad Sci U S A, 2006. **103**(2): p. 395-400.
235. Larijani, M., et al., *Methylation protects cytidines from AID-mediated deamination.* Mol Immunol, 2005. **42**(5): p. 599-604.
236. Nabel, C.S., et al., *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation.* Nature chemical biology, 2012. **8**(9): p. 751-8.
237. Morgan, H.D., et al., *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming.* The Journal of biological chemistry, 2004. **279**(50): p. 52353-60.
238. Bhutani, N., et al., *Reprogramming towards pluripotency requires AID-dependent DNA demethylation.* Nature, 2010. **463**(7284): p. 1042-7.
239. Vuong, B.Q. and J. Chaudhuri, *Combinatorial mechanisms regulating AID-dependent DNA deamination: interacting proteins and post-translational modifications.* Seminars in immunology, 2012. **24**(4): p. 264-72.
240. Casellas, R., et al., *Ku80 is required for immunoglobulin isotype switching.* The EMBO journal, 1998. **17**(8): p. 2404-11.
241. Manis, J.P., et al., *Ku70 is required for late B cell development and immunoglobulin heavy chain class switching.* The Journal of experimental medicine, 1998. **187**(12): p. 2081-9.
242. Kotnis, A., et al., *Non-homologous end joining in class switch recombination: the beginning of the end.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2009. **364**(1517): p. 653-65.
243. Xu, Z., et al., *Immunoglobulin class-switch DNA recombination: induction, targeting and beyond.* Nature reviews. Immunology, 2012. **12**(7): p. 517-31.
244. Imai, K., et al., *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nat Immunol, 2003. **4**(10): p. 1023-8.
245. Schrader, C.E., J. Vardo, and J. Stavnezer, *Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions.* The Journal of experimental medicine, 2002. **195**(3): p. 367-73.
246. Soulas-Sprauel, P., et al., *Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination.* The Journal of experimental medicine, 2007. **204**(7): p. 1717-27.
247. Stavnezer, J. and C.E. Schrader, *Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination.* Trends Genet, 2006. **22**(1): p. 23-8.
248. Aranda, S., G. Mas, and L. Di Croce, *Regulation of gene transcription by Polycomb proteins.* Sci Adv, 2015. **1**(11): p. e1500737.
249. Trojer, P. and D. Reinberg, *Facultative heterochromatin: is there a distinctive molecular signature?* Mol Cell, 2007. **28**(1): p. 1-13.
250. Fabian, M.R., N. Sonenberg, and W. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* Annu Rev Biochem, 2010. **79**: p. 351-79.

251. Khraiwesh, B., et al., *Transcriptional control of gene expression by microRNAs.* Cell, 2010. **140**(1): p. 111-22.
252. Kung, J.T., D. Colognori, and J.T. Lee, *Long noncoding RNAs: past, present, and future.* Genetics, 2013. **193**(3): p. 651-69.
253. Mercer, T.R. and J.S. Mattick, *Structure and function of long noncoding RNAs in epigenetic regulation.* Nat Struct Mol Biol, 2013. **20**(3): p. 300-7.
254. Kalashnikova, A.A., R.A. Rogge, and J.C. Hansen, *Linker histone H1 and protein-protein interactions.* Biochim Biophys Acta, 2016. **1859**(3): p. 455-61.
255. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.
256. Hassan, A.H., et al., *Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes.* Cell, 2002. **111**(3): p. 369-79.
257. Kim, J., et al., *Tudor, MBT and chromo domains gauge the degree of lysine methylation.* EMBO Rep, 2006. **7**(4): p. 397-403.
258. Hon, G.C., R.D. Hawkins, and B. Ren, *Predictive chromatin signatures in the mammalian genome.* Hum Mol Genet, 2009. **18**(R2): p. R195-201.
259. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
260. Schneider, R., et al., *Histone H3 lysine 4 methylation patterns in higher eukaryotic genes.* Nat Cell Biol, 2004. **6**(1): p. 73-7.
261. Sims, R.J., 3rd, et al., *Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains.* J Biol Chem, 2005. **280**(51): p. 41789-92.
262. Zegerman, P., et al., *Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex.* J Biol Chem, 2002. **277**(14): p. 11621-4.
263. Bannister, A.J., et al., *Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes.* J Biol Chem, 2005. **280**(18): p. 17732-6.
264. Juan, A.H., et al., *Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells.* Cell Rep, 2016. **17**(5): p. 1369-1382.
265. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.
266. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.
267. Lehnertz, B., et al., *Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin.* Curr Biol, 2003. **13**(14): p. 1192-200.
268. Bartke, T., et al., *Nucleosome-interacting proteins regulated by DNA and histone methylation.* Cell, 2010. **143**(3): p. 470-84.
269. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
270. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression.* Trends Genet, 2016. **32**(1): p. 42-56.
271. Gold, M. and J. Hurwitz, *The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. V. Purification and Properties of the Deoxyribonucleic Acid-Methylating Activity of Escherichia Coli.* J Biol Chem, 1964. **239**: p. 3858-65.
272. Hurwitz, J., M. Gold, and M. Anders, *The Enzymatic Methylation of Ribonucleic Acid and Deoxyribonucleic Acid. Iv. The Properties of the Soluble Ribonucleic Acid-Methylating Enzymes.* J Biol Chem, 1964. **239**: p. 3474-82.
273. Arber, W. and D. Dussoix, *Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda.* J Mol Biol, 1962. **5**: p. 18-36.

274. Dussoix, D. and W. Arber, *Host specificity of DNA produced by Escherichia coli. II. Control over acceptance of DNA from infecting phage lambda.* J Mol Biol, 1962. **5**: p. 37-49.
275. Arber, W., *Host Specificity of DNA Produced by Escherichia Coli V . The Role of Methionine in the Production of Host Specificity.* J Mol Biol, 1965. **11**: p. 247-56.
276. Arber, W. and M.L. Morse, *Host Specificity of DNA Produced by Escherichia Coli. Vi. Effects on Bacterial Conjugation.* Genetics, 1965. **51**: p. 137-48.
277. Dussoix, D. and W. Arber, *Host Specificity of DNA Produced by Escherichia Coli. Iv. Host Specificity of Infectious DNA from Bacteriophage Lambda.* J Mol Biol, 1965. **11**: p. 238-46.
278. Kelly, T.J., Jr. and H.O. Smith, *A restriction enzyme from Hemophilus influenzae. II.* J Mol Biol, 1970. **51**(2): p. 393-409.
279. Smith, H.O. and K.W. Wilcox, *A restriction enzyme from Hemophilus influenzae. I. Purification and general properties.* J Mol Biol, 1970. **51**(2): p. 379-91.
280. Adler, S.P. and D. Nathans, *Studies of SV 40 DNA. V. Conversion of circular to linear SV 40 DNA by restriction endonuclease from Escherichia coli B.* Biochim Biophys Acta, 1973. **299**(2): p. 177-88.
281. Danna, K. and D. Nathans, *Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae.* Proc Natl Acad Sci U S A, 1971. **68**(12): p. 2913-7.
282. Danna, K.J., G.H. Sack, Jr., and D. Nathans, *Studies of simian virus 40 DNA. VII. A cleavage map of the SV40 genome.* J Mol Biol, 1973. **78**(2): p. 363-76.
283. Khoury, G., et al., *Characterization of a rearrangement in viral DNA: mapping of the circular simian virus 40-like DNA containing a triplication of a specific one-third of the viral genome.* J Mol Biol, 1974. **87**(2): p. 289-301.
284. Martin, M.A., et al., *Characterization of "heavy" and "light" SV40-like particles from a patient with PML.* Virology, 1974. **59**(1): p. 179-89.
285. Sack, G.H., Jr. and D. Nathans, *Studies of SV40 DNA. VI. Cleavage of SV40 DNA by restriction endonuclease from Hemophilus parainfluenzae.* Virology, 1973. **51**(2): p. 517-20.
286. Glickman, B.W. and M. Radman, *Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction.* Proc Natl Acad Sci U S A, 1980. **77**(2): p. 1063-7.
287. Marinus, M.G., *Adenine methylation of Okazaki fragments in Escherichia coli.* J Bacteriol, 1976. **128**(3): p. 853-4.
288. Marinus, M.G. and E.B. Konrad, *Hyper-recombination in dam mutants of Escherichia coli K-12.* Mol Gen Genet, 1976. **149**(3): p. 273-7.
289. Radman, M., et al., *Replicational fidelity: mechanisms of mutation avoidance and mutation fixation.* Cold Spring Harb Symp Quant Biol, 1979. **43 Pt 2**: p. 937-46.
290. Marinus, M.G. and N.R. Morris, *Biological function for 6-methyladenine residues in the DNA of Escherichia coli K12.* J Mol Biol, 1974. **85**(2): p. 309-22.
291. Marinus, M.G., A. Poteete, and J.A. Arraj, *Correlation of DNA adenine methylase activity with spontaneous mutability in Escherichia coli K-12.* Gene, 1984. **28**(1): p. 123-5.
292. Schlagman, S.L., S. Hattman, and M.G. Marinus, *Direct role of the Escherichia coli Dam DNA methyltransferase in methylation-directed mismatch repair.* J Bacteriol, 1986. **165**(3): p. 896-900.
293. Herman, G.E. and P. Modrich, *Escherichia coli K-12 clones that overproduce dam methylase are hypermutable.* J Bacteriol, 1981. **145**(1): p. 644-6.
294. Rydberg, B., *Bromouracil mutagenesis and mismatch repair in mutator strains of Escherichia coli.* Mutat Res, 1978. **52**(1): p. 11-24.
295. Bauer, J., G. Krammer, and R. Knippers, *Asymmetric repair of bacteriophage T7 heteroduplex DNA.* Mol Gen Genet, 1981. **181**(4): p. 541-7.
296. Sheid, B., P.R. Srinivasan, and E. Borek, *Deoxyribonucleic acid methylase of mammalian tissues.* Biochemistry, 1968. **7**(1): p. 280-5.
297. Kalousek, F. and N.R. Morris, *Deoxyribonucleic acid methylase activity in rat spleen.* J Biol Chem, 1968. **243**(9): p. 2440-2.

298. Gruenbaum, Y., et al., *Sequence specificity of methylation in higher plant DNA.* Nature, 1981. **292**(5826): p. 860-2.
299. Hardcastle, T.J., *High-throughput sequencing of cytosine methylation in plant DNA.* Plant Methods, 2013. **9**(1): p. 16.
300. Gruenbaum, Y., et al., *Methylation of CpG sequences in eukaryotic DNA.* FEBS Lett, 1981. **124**(1): p. 67-71.
301. Naveh-Many, T. and H. Cedar, *Active gene sequences are undermethylated.* Proc Natl Acad Sci U S A, 1981. **78**(7): p. 4246-50.
302. Swartz, M.N., T.A. Trautner, and A. Kornberg, *Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids.* J Biol Chem, 1962. **237**: p. 1961-7.
303. Coulondre, C., et al., *Molecular basis of base substitution hotspots in Escherichia coli.* Nature, 1978. **274**(5673): p. 775-80.
304. Bird, A.P., *DNA methylation and the frequency of CpG in animal DNA.* Nucleic Acids Res, 1980. **8**(7): p. 1499-504.
305. Sved, J. and A. Bird, *The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4692-6.
306. Cooper, D.N. and H. Youssoufian, *The CpG dinucleotide and human genetic disease.* Hum Genet, 1988. **78**(2): p. 151-5.
307. Bulmer, M., *Neighboring base effects on substitution rates in pseudogenes.* Mol Biol Evol, 1986. **3**(4): p. 322-9.
308. Cooper, D.N., M.H. Taggart, and A.P. Bird, *Unmethylated domains in vertebrate DNA.* Nucleic Acids Res, 1983. **11**(3): p. 647-58.
309. McClelland, M. and R. Ivarie, *Asymmetrical distribution of CpG in an 'average' mammalian gene.* Nucleic Acids Res, 1982. **10**(23): p. 7865-77.
310. Stein, R., et al., *Pattern of methylation of two genes coding for housekeeping functions.* Proc Natl Acad Sci U S A, 1983. **80**(9): p. 2422-6.
311. Busslinger, M., J. Hurst, and R.A. Flavell, *DNA methylation and the regulation of globin gene expression.* Cell, 1983. **34**(1): p. 197-206.
312. Liskay, R.M. and R.J. Evans, *Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyltransferase gene.* Proc Natl Acad Sci U S A, 1980. **77**(8): p. 4895-8.
313. Venolia, L., et al., *Transformation with DNA from 5-azacytidine-reactivated X chromosomes.* Proc Natl Acad Sci U S A, 1982. **79**(7): p. 2352-4.
314. Yen, P.H., et al., *Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes.* Proc Natl Acad Sci U S A, 1984. **81**(6): p. 1759-63.
315. Tykocinski, M.L. and E.E. Max, *CG dinucleotide clusters in MHC genes and in 5' demethylated genes.* Nucleic Acids Res, 1984. **12**(10): p. 4385-96.
316. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.
317. Riggs, A.D. and G.P. Pfeifer, *X-chromosome inactivation and cell memory.* Trends Genet, 1992. **8**(5): p. 169-74.
318. Neumann, B. and D.P. Barlow, *Multiple roles for DNA methylation in gametic imprinting.* Curr Opin Genet Dev, 1996. **6**(2): p. 159-63.
319. Razin, A. and H. Cedar, *DNA methylation and genomic imprinting.* Cell, 1994. **77**(4): p. 473-6.
320. Bird, A.P., *The relationship of DNA methylation to cancer.* Cancer Surv, 1996. **28**: p. 87-101.
321. Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation.* N Engl J Med, 2003. **349**(21): p. 2042-54.
322. Esteller, M., et al., *MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas.* Oncogene, 1998. **17**(18): p. 2413-7.

323. Kanaya, T., et al., *Frequent hypermethylation of MLH1 promoter in normal endometrium of patients with endometrial cancers.* Oncogene, 2003. **22**(15): p. 2352-60.
324. Li, X., et al., *MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features.* PLoS One, 2013. **8**(3): p. e59064.
325. Simpkins, S.B., et al., *MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers.* Hum Mol Genet, 1999. **8**(4): p. 661-6.
326. Okano, M., S. Xie, and E. Li, *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases.* Nat Genet, 1998. **19**(3): p. 219-20.
327. Wu, H., et al., *Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells.* Genes Dev, 2011. **25**(7): p. 679-84.
328. Ficz, G., et al., *Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.* Nature, 2011. **473**(7347): p. 398-402.
329. Wu, H., et al., *Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells.* Nature, 2011. **473**(7347): p. 389-93.
330. Ben-Hattar, J. and J. Jiricny, *Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo.* Gene, 1988. **65**(2): p. 219-27.
331. Saluz, H.P., et al., *Genomic sequencing and in vivo footprinting of an expression-specific DNase I-hypersensitive site of avian vitellogenin II promoter reveal a demethylation of a mCpG and a change in specific interactions of proteins with DNA.* Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6697-700.
332. Saluz, H.P., J. Jiricny, and J.P. Jost, *Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(19): p. 7167-71.
333. Keshet, I., J. Lieman-Hurwitz, and H. Cedar, *DNA methylation affects the formation of active chromatin.* Cell, 1986. **44**(4): p. 535-43.
334. Becker, P.B., S. Ruppert, and G. Schutz, *Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors.* Cell, 1987. **51**(3): p. 435-43.
335. Iguchi-Ariga, S.M. and W. Schaffner, *CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation.* Genes Dev, 1989. **3**(5): p. 612-9.
336. Medvedeva, Y.A., et al., *Effects of cytosine methylation on transcription factor binding sites.* BMC Genomics, 2014. **15**: p. 119.
337. Comb, M. and H.M. Goodman, *CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2.* Nucleic Acids Res, 1990. **18**(13): p. 3975-82.
338. Falzon, M. and E.L. Kuff, *Binding of the transcription factor EBP-80 mediates the methylation response of an intracisternal A-particle long terminal repeat promoter.* Mol Cell Biol, 1991. **11**(1): p. 117-25.
339. Gaston, K. and M. Fried, *CpG methylation and the binding of YY1 and ETS proteins to the Surf-1/Surf-2 bidirectional promoter.* Gene, 1995. **157**(1-2): p. 257-9.
340. Gaston, K. and M. Fried, *CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes.* Nucleic Acids Res, 1995. **23**(6): p. 901-9.
341. Radtke, F., et al., *Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites.* Biol Chem Hoppe Seyler, 1996. **377**(1): p. 47-56.
342. Meehan, R.R., et al., *Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs.* Cell, 1989. **58**(3): p. 499-507.

343. Boyes, J. and A. Bird, *DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein.* Cell, 1991. **64**(6): p. 1123-34.
344. Lewis, J.D., et al., *Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA.* Cell, 1992. **69**(6): p. 905-14.
345. Nan, X., et al., *DNA methylation specifies chromosomal localization of MeCP2.* Mol Cell Biol, 1996. **16**(1): p. 414-21.
346. Tate, P., W. Skarnes, and A. Bird, *The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse.* Nat Genet, 1996. **12**(2): p. 205-8.
347. Nan, X., F.J. Campoy, and A. Bird, *MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin.* Cell, 1997. **88**(4): p. 471-81.
348. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. **393**(6683): p. 386-9.
349. Ng, H.H., et al., *MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex.* Nat Genet, 1999. **23**(1): p. 58-61.
350. Hendrich, B., et al., *Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.* Genes Dev, 2001. **15**(6): p. 710-23.
351. Martin Caballero, I., et al., *The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation.* PLoS One, 2009. **4**(1): p. e4315.
352. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.* Nat Genet, 1999. **23**(2): p. 185-8.
353. Moretti, P., et al., *Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome.* J Neurosci, 2006. **26**(1): p. 319-27.
354. Moretti, P. and H.Y. Zoghbi, *MeCP2 dysfunction in Rett syndrome and related disorders.* Curr Opin Genet Dev, 2006. **16**(3): p. 276-81.
355. Guy, J., et al., *Reversal of neurological defects in a mouse model of Rett syndrome.* Science, 2007. **315**(5815): p. 1143-7.
356. Struhl, K., *Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins.* Trends Biochem Sci, 1989. **14**(4): p. 137-40.
357. Busch, S.J. and P. Sassone-Corsi, *Dimers, leucine zippers and DNA-binding domains.* Trends Genet, 1990. **6**(2): p. 36-40.
358. Abel, T. and T. Maniatis, *Gene regulation. Action of leucine zippers.* Nature, 1989. **341**(6237): p. 24-5.
359. Luscher, B. and L.G. Larsson, *The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation.* Oncogene, 1999. **18**(19): p. 2955-66.
360. Boyes, J. and A. Bird, *Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein.* EMBO J, 1992. **11**(1): p. 327-33.
361. Hsieh, C.L., *Dependence of transcriptional repression on CpG methylation density.* Mol Cell Biol, 1994. **14**(8): p. 5487-94.
362. Kass, S.U., N. Landsberger, and A.P. Wolffe, *DNA methylation directs a time-dependent repression of transcription initiation.* Curr Biol, 1997. **7**(3): p. 157-65.
363. Kass, S.U., D. Pruss, and A.P. Wolffe, *How does DNA methylation repress transcription?* Trends Genet, 1997. **13**(11): p. 444-9.
364. Stringer, K.F., C.J. Ingles, and J. Greenblatt, *Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID.* Nature, 1990. **345**(6278): p. 783-6.
365. Courey, A.J. and R. Tjian, *Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif.* Cell, 1988. **55**(5): p. 887-98.
366. Dynan, W.S. and R. Tjian, *The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter.* Cell, 1983. **35**(1): p. 79-87.

367. Emili, A., J. Greenblatt, and C.J. Ingles, *Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein.* Mol Cell Biol, 1994. **14**(3): p. 1582-93.
368. Sentenac, A., *Eukaryotic RNA polymerases.* CRC Crit Rev Biochem, 1985. **18**(1): p. 31-90.
369. Turowski, T.W. and D. Tollervey, *Transcription by RNA polymerase III: insights into mechanism and regulation.* Biochem Soc Trans, 2016. **44**(5): p. 1367-1375.
370. Allison, L.A., et al., *The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function.* Mol Cell Biol, 1988. **8**(1): p. 321-9.
371. Hsin, J.P. and J.L. Manley, *The RNA polymerase II CTD coordinates transcription and RNA processing.* Genes Dev, 2012. **26**(19): p. 2119-37.
372. Lee, T.I. and R.A. Young, *Transcription of eukaryotic protein-coding genes.* Annu Rev Genet, 2000. **34**: p. 77-137.
373. Hermann, A., S. Schmitt, and A. Jeltsch, *The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity.* J Biol Chem, 2003. **278**(34): p. 31717-21.
374. Aapola, U., et al., *Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family.* Genomics, 2000. **65**(3): p. 293-8.
375. Chedin, F., M.R. Lieber, and C.L. Hsieh, *The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16916-21.
376. Gowher, H., et al., *Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L.* J Biol Chem, 2005. **280**(14): p. 13341-8.
377. Santi, D.V., C.E. Garrett, and P.J. Barr, *On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs.* Cell, 1983. **33**(1): p. 9-10.
378. Santi, D.V., A. Norment, and C.E. Garrett, *Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine.* Proc Natl Acad Sci U S A, 1984. **81**(22): p. 6993-7.
379. Wu, J.C. and D.V. Santi, *Kinetic and catalytic mechanism of HhaI methyltransferase.* J Biol Chem, 1987. **262**(10): p. 4778-86.
380. Chen, L., et al., *Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase.* Biochemistry, 1991. **30**(46): p. 11018-25.
381. Smith, S.S., et al., *Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4744-8.
382. Klimasauskas, S., et al., *HhaI methyltransferase flips its target base out of the DNA helix.* Cell, 1994. **76**(2): p. 357-69.
383. Bestor, T., et al., *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases.* J Mol Biol, 1988. **203**(4): p. 971-83.
384. Yoder, J.A., et al., *DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe.* J Mol Biol, 1997. **270**(3): p. 385-95.
385. Arita, K., et al., *Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism.* Nature, 2008. **455**(7214): p. 818-21.
386. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.
387. Dhe-Paganon, S., F. Syeda, and L. Park, *DNA methyl transferase 1: regulatory mechanisms and implications in health and disease.* Int J Biochem Mol Biol, 2011. **2**(1): p. 58-66.
388. Schermelleh, L., et al., *Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation.* Nucleic Acids Res, 2007. **35**(13): p. 4301-12.
389. Spada, F., et al., *DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells.* J Cell Biol, 2007. **176**(5): p. 565-71.

390. Bestor, T.H., *Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain.* EMBO J, 1992. **11**(7): p. 2611-7.
391. Bacolla, A., et al., *Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna.* J Biol Chem, 1999. **274**(46): p. 33011-9.
392. Pradhan, S., et al., *Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation.* J Biol Chem, 1999. **274**(46): p. 33002-10.
393. Lei, H., et al., *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.* Development, 1996. **122**(10): p. 3195-205.
394. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
395. Hsieh, C.L., *In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b.* Mol Cell Biol, 1999. **19**(12): p. 8211-8.
396. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
397. Miniou, P., et al., *Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients.* Hum Mol Genet, 1994. **3**(12): p. 2093-102.
398. Hansen, R.S., et al., *Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant.* Hum Mol Genet, 2000. **9**(18): p. 2575-87.
399. Kondo, T., et al., *Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2.* Hum Mol Genet, 2000. **9**(4): p. 597-604.
400. Xu, G.L., et al., *Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene.* Nature, 1999. **402**(6758): p. 187-91.
401. Wijmenga, C., et al., *Genetic variation in ICF syndrome: evidence for genetic heterogeneity.* Hum Mutat, 2000. **16**(6): p. 509-17.
402. Ehrlich, M., *The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease.* Clin Immunol, 2003. **109**(1): p. 17-28.
403. Jin, B. and K.D. Robertson, *DNA methyltransferases, DNA damage repair, and cancer.* Adv Exp Med Biol, 2013. **754**: p. 3-29.
404. Subramaniam, D., et al., *DNA methyltransferases: a novel target for prevention and therapy.* Front Oncol, 2014. **4**: p. 80.
405. Issa, J.P., *Decitabine.* Curr Opin Oncol, 2003. **15**(6): p. 446-51.
406. Issa, J.P., et al., *Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies.* Blood, 2004. **103**(5): p. 1635-40.
407. Kantarjian, H.M., et al., *Results of decitabine (5-aza-2'deoxycytidine) therapy in 130 patients with chronic myelogenous leukemia.* Cancer, 2003. **98**(3): p. 522-8.
408. Momparler, R.L. and J. Ayoub, *Potential of 5-aza-2'-deoxycytidine (Decitabine) a potent inhibitor of DNA methylation for therapy of advanced non-small cell lung cancer.* Lung Cancer, 2001. **34 Suppl 4**: p. S111-5.
409. Mayer, W., et al., *Demethylation of the zygotic paternal genome.* Nature, 2000. **403**(6769): p. 501-2.
410. Oswald, J., et al., *Active demethylation of the paternal genome in the mouse zygote.* Current biology : CB, 2000. **10**(8): p. 475-8.
411. Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mechanisms of development, 2002. **117**(1-2): p. 15-23.
412. Seisenberger, S., et al., *The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.* Mol Cell, 2012. **48**(6): p. 849-62.
413. Zhu, J.K., *Active DNA demethylation mediated by DNA glycosylases.* Annu Rev Genet, 2009. **43**: p. 143-66.

414. Jost, J.P., et al., *Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase.* J Biol Chem, 1995. **270**(17): p. 9734-9.
415. Vairapandi, M. and N.J. Duker, *Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase.* Nucleic Acids Res, 1993. **21**(23): p. 5323-7.
416. Popp, C., et al., *Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.* Nature, 2010. **463**(7284): p. 1101-5.
417. Santos, F., et al., *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair.* Epigenetics Chromatin, 2013. **6**(1): p. 39.
418. Lorsbach, R.B., et al., *TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23).* Leukemia, 2003. **17**(3): p. 637-41.
419. Ono, R., et al., *LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23).* Cancer Res, 2002. **62**(14): p. 4075-80.
420. Maiti, A. and A.C. Drohat, *Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites.* J Biol Chem, 2011. **286**(41): p. 35334-8.
421. Szwagierczak, A., et al., *Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA.* Nucleic Acids Res, 2010. **38**(19): p. e181.
422. Globisch, D., et al., *Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.* PLoS One, 2010. **5**(12): p. e15367.
423. Kriaucionis, S. and N. Heintz, *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.* Science, 2009. **324**(5929): p. 929-30.
424. Pastor, W.A., L. Aravind, and A. Rao, *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.* Nat Rev Mol Cell Biol, 2013. **14**(6): p. 341-56.
425. Loenarz, C. and C.J. Schofield, *Expanding chemical biology of 2-oxoglutarate oxygenases.* Nat Chem Biol, 2008. **4**(3): p. 152-6.
426. Loenarz, C. and C.J. Schofield, *Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases.* Trends Biochem Sci, 2011. **36**(1): p. 7-18.
427. Ponnaluri, V.K., J.P. Maciejewski, and M. Mukherji, *A mechanistic overview of TET-mediated 5-methylcytosine oxidation.* Biochem Biophys Res Commun, 2013. **436**(2): p. 115-20.
428. Kovaleva, E.G. and J.D. Lipscomb, *Versatility of biological non-heme Fe(II) centers in oxygen activation reactions.* Nat Chem Biol, 2008. **4**(3): p. 186-93.
429. Minor, E.A., et al., *Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine.* J Biol Chem, 2013. **288**(19): p. 13669-74.
430. Costas, M., et al., *Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates.* Chem Rev, 2004. **104**(2): p. 939-86.
431. Ozer, A. and R.K. Bruick, *Non-heme dioxygenases: cellular sensors and regulators jelly rolled into one?* Nat Chem Biol, 2007. **3**(3): p. 144-53.
432. Hashimoto, H., et al., *Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA.* Nucleic Acids Res, 2015. **43**(22): p. 10713-21.
433. Hashimoto, H., et al., *Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA.* Nature, 2014. **506**(7488): p. 391-5.
434. Hu, L., et al., *Structural insight into substrate preference for TET-mediated oxidation.* Nature, 2015. **527**(7576): p. 118-22.
435. Allen, M.D., et al., *Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase.* EMBO J, 2006. **25**(19): p. 4503-12.
436. Lee, J.H., K.S. Voo, and D.G. Skalnik, *Identification and characterization of the DNA binding domain of CpG-binding protein.* J Biol Chem, 2001. **276**(48): p. 44669-76.

437. Hu, L., et al., *Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation.* Cell, 2013. **155**(7): p. 1545-55.
438. Iqbal, K., et al., *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3642-7.
439. Wossidlo, M., et al., *5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming.* Nat Commun, 2011. **2**: p. 241.
440. Williams, K., et al., *TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.* Nature, 2011. **473**(7347): p. 343-8.
441. Pastor, W.A., et al., *Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells.* Nature, 2011. **473**(7347): p. 394-7.
442. Dawlaty, M.M., et al., *Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development.* Cell Stem Cell, 2011. **9**(2): p. 166-75.
443. Zhao, Z., et al., *Combined Loss of Tet1 and Tet2 Promotes B Cell, but Not Myeloid Malignancies, in Mice.* Cell Rep, 2015. **13**(8): p. 1692-704.
444. Hackett, J.A., et al., *Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.* Science, 2013. **339**(6118): p. 448-52.
445. Vincent, J.J., et al., *Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells.* Cell Stem Cell, 2013. **12**(4): p. 470-8.
446. Ko, M., et al., *Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2.* Nature, 2010. **468**(7325): p. 839-43.
447. Langemeijer, S.M., et al., *Acquired mutations in TET2 are common in myelodysplastic syndromes.* Nat Genet, 2009. **41**(7): p. 838-42.
448. Smith, A.E., et al., *Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value.* Blood, 2010. **116**(19): p. 3923-32.
449. Li, Z., et al., *Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies.* Blood, 2011. **118**(17): p. 4509-18.
450. Moran-Crusio, K., et al., *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation.* Cancer Cell, 2011. **20**(1): p. 11-24.
451. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.* Nature, 2010. **466**(7310): p. 1129-33.
452. Gu, T.P., et al., *The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes.* Nature, 2011. **477**(7366): p. 606-10.
453. Amouroux, R., et al., *De novo DNA methylation drives 5hmC accumulation in mouse zygotes.* Nat Cell Biol, 2016. **18**(2): p. 225-33.
454. Inoue, A., et al., *Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development.* Cell Res, 2011. **21**(12): p. 1670-6.
455. Inoue, A. and Y. Zhang, *Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos.* Science, 2011. **334**(6053): p. 194.
456. Valinluck, V. and L.C. Sowers, *Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1.* Cancer Res, 2007. **67**(3): p. 946-50.
457. Hashimoto, H., et al., *Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation.* Nucleic Acids Res, 2012. **40**(11): p. 4841-9.
458. Chen, D., et al., *T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha.* J Biol Chem, 2003. **278**(40): p. 38586-92.
459. Jost, J.P., S. Thiry, and M. Siegmann, *Estradiol receptor potentiates, in vitro, the activity of 5-methylcytosine DNA glycosylase.* FEBS Lett, 2002. **527**(1-3): p. 63-6.
460. Zhu, B., et al., *Overexpression of 5-methylcytosine DNA glycosylase in human embryonic kidney cells EcR293 demethylates the promoter of a hormone-regulated reporter gene.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5031-6.

461. Tini, M., et al., *Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription.* Mol Cell, 2002. **9**(2): p. 265-77.
462. Tanaka, Y., et al., *Extensive brain hemorrhage and embryonic lethality in a mouse null mutant of CREB-binding protein.* Mech Dev, 2000. **95**(1-2): p. 133-45.
463. Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.
464. Jost, J.P., et al., *5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation.* Nucleic Acids Res, 2001. **29**(21): p. 4452-61.
465. Hu, X.V., et al., *Identification of RING finger protein 4 (RNF4) as a modulator of DNA demethylation through a functional genomics screen.* Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15087-92.
466. Wang, Y., *RING finger protein 4 (RNF4) derepresses gene expression from DNA methylation.* J Biol Chem, 2014. **289**(49): p. 33808-13.
467. Xu, J., et al., *Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells.* Proc Natl Acad Sci U S A, 2007. **104**(30): p. 12377-82.
468. Thomassin, H., et al., *Glucocorticoid-induced DNA demethylation and gene memory during development.* EMBO J, 2001. **20**(8): p. 1974-83.
469. Barreto, G., et al., *Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation.* Nature, 2007. **445**(7128): p. 671-5.
470. Jin, S.G., C. Guo, and G.P. Pfeifer, *GADD45A does not promote DNA demethylation.* PLoS Genet, 2008. **4**(3): p. e1000013.
471. Shen, L., et al., *Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics.* Cell, 2013. **153**(3): p. 692-706.
472. Song, C.X., et al., *Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming.* Cell, 2013. **153**(3): p. 678-91.
473. Kangaspeska, S., et al., *Transient cyclical methylation of promoter DNA.* Nature, 2008. **452**(7183): p. 112-5.
474. Metivier, R., et al., *Cyclical DNA methylation of a transcriptionally active promoter.* Nature, 2008. **452**(7183): p. 45-50.
475. Di Croce, L., et al., *Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor.* Science, 2002. **295**(5557): p. 1079-82.
476. Yi, Y.W., et al., *Gadd45 family proteins are coactivators of nuclear hormone receptors.* Biochem Biophys Res Commun, 2000. **272**(1): p. 193-8.
477. Berg, J.M., *DNA binding specificity of steroid receptors.* Cell, 1989. **57**(7): p. 1065-8.
478. Klug, A. and J.W. Schwabe, *Protein motifs 5. Zinc fingers.* FASEB J, 1995. **9**(8): p. 597-604.
479. Beato, M., *Transcriptional control by nuclear receptors.* FASEB J, 1991. **5**(7): p. 2044-51.
480. Perlmann, T., et al., *Determinants for selective RAR and TR recognition of direct repeat HREs.* Genes Dev, 1993. **7**(7B): p. 1411-22.
481. Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-9.
482. Wurtz, J.M., et al., *A canonical structure for the ligand-binding domain of nuclear receptors.* Nat Struct Biol, 1996. **3**(1): p. 87-94.
483. Warnmark, A., et al., *Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation.* Mol Endocrinol, 2003. **17**(10): p. 1901-9.
484. Pratt, W.B. and D.O. Toft, *Steroid receptor interactions with heat shock protein and immunophilin chaperones.* Endocr Rev, 1997. **18**(3): p. 306-60.
485. Xu, J. and Q. Li, *Review of the in vivo functions of the p160 steroid receptor coactivator family.* Mol Endocrinol, 2003. **17**(9): p. 1681-92.

486. Grimaldi, M., et al., *Reporter Cell Lines for the Characterization of the Interactions between Human Nuclear Receptors and Endocrine Disruptors.* Front Endocrinol (Lausanne), 2015. **6**: p. 62.
487. Green, S., et al., *Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A.* Nature, 1986. **320**(6058): p. 134-9.
488. Mosselman, S., J. Polman, and R. Dijkema, *ER beta: identification and characterization of a novel human estrogen receptor.* FEBS Lett, 1996. **392**(1): p. 49-53.
489. Klein-Hitpass, L., et al., *A 13 bp palindrome is a functional estrogen responsive element and interacts specifically with estrogen receptor.* Nucleic Acids Res, 1988. **16**(2): p. 647-63.
490. Enmark, E., et al., *Human estrogen receptor beta-gene structure, chromosomal localization, and expression pattern.* J Clin Endocrinol Metab, 1997. **82**(12): p. 4258-65.
491. Kuiper, G.G., et al., *Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta.* Endocrinology, 1997. **138**(3): p. 863-70.
492. Katzenellenbogen, B.S. and J.A. Katzenellenbogen, *Estrogen receptor transcription and transactivation: Estrogen receptor alpha and estrogen receptor beta: regulation by selective estrogen receptor modulators and importance in breast cancer.* Breast Cancer Res, 2000. **2**(5): p. 335-44.
493. Barkhem, T., et al., *Differential response of estrogen receptor alpha and estrogen receptor beta to partial estrogen agonists/antagonists.* Mol Pharmacol, 1998. **54**(1): p. 105-12.
494. Kuiper, G.G., et al., *Cloning of a novel receptor expressed in rat prostate and ovary.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 5925-30.
495. Tremblay, G.B., et al., *Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor beta.* Mol Endocrinol, 1997. **11**(3): p. 353-65.
496. Pfaffl, M.W., et al., *Tissue-specific expression pattern of estrogen receptors (ER): quantification of ER alpha and ER beta mRNA with real-time RT-PCR.* APMIS, 2001. **109**(5): p. 345-55.
497. Klein-Hitpass, L., et al., *Specific binding of estrogen receptor to the estrogen response element.* Mol Cell Biol, 1989. **9**(1): p. 43-9.
498. Joshi, S.R., R.B. Ghattamaneni, and W.M. Scovell, *Expanding the paradigm for estrogen receptor binding and transcriptional activation.* Mol Endocrinol, 2011. **25**(6): p. 980-94.
499. Fullwood, M.J., et al., *An oestrogen-receptor-alpha-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.
500. Kraus, W.L. and J.T. Kadonaga, *p300 and estrogen receptor cooperatively activate transcription via differential enhancement of initiation and reinitiation.* Genes Dev, 1998. **12**(3): p. 331-42.
501. Duffy, M.J., *Estrogen receptors: role in breast cancer.* Crit Rev Clin Lab Sci, 2006. **43**(4): p. 325-47.
502. Ali, S. and R.C. Coombes, *Estrogen receptor alpha in human breast cancer: occurrence and significance.* J Mammary Gland Biol Neoplasia, 2000. **5**(3): p. 271-81.
503. Wilks, A.F., et al., *Estrogen induces a demethylation at the 5' end region of the chicken vitellogenin gene.* Proc Natl Acad Sci U S A, 1982. **79**(14): p. 4252-5.
504. Meijlink, F.C., et al., *Methylation of the chicken vitellogenin gene: influence of estradiol administration.* Nucleic Acids Res, 1983. **11**(5): p. 1361-73.
505. Burch, J.B. and H. Weintraub, *Temporal order of chromatin structural changes associated with activation of the major chicken vitellogenin gene.* Cell, 1983. **33**(1): p. 65-76.
506. Wilks, A., M. Seldran, and J.P. Jost, *An estrogen-dependent demethylation at the 5' end of the chicken vitellogenin gene is independent of DNA synthesis.* Nucleic Acids Res, 1984. **12**(2): p. 1163-77.
507. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA.* Proc Natl Acad Sci U S A, 1977. **74**(2): p. 560-4.
508. Church, G.M. and W. Gilbert, *Genomic sequencing.* Proc Natl Acad Sci U S A, 1984. **81**(7): p. 1991-5.

509. Feavers, I.M., et al., *Interaction of two nonhistone proteins with the estradiol response element of the avian vitellogenin gene modulates the binding of estradiol-receptor complex.* Proc Natl Acad Sci U S A, 1987. **84**(21): p. 7453-7.
510. Burch, J.B., et al., *Two functional estrogen response elements are located upstream of the major chicken vitellogenin gene.* Mol Cell Biol, 1988. **8**(3): p. 1123-31.
511. Hoeijmakers, J.H., *DNA damage, aging, and cancer.* N Engl J Med, 2009. **361**(15): p. 1475-85.
512. O'Driscoll, M., *Diseases associated with defective responses to DNA damage.* Cold Spring Harb Perspect Biol, 2012. **4**(12).
513. Pekarik, V., et al., *Screening for gene function in chicken embryo using RNAi and electroporation.* Nature biotechnology, 2003. **21**(1): p. 93-6.
514. Baerenfaller, K., F. Fischer, and J. Jiricny, *Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro.* Methods in enzymology, 2006. **408**: p. 285-303.
515. Baeriswyl, T., O. Mauti, and E.T. Stoeckli, *Temporal control of gene silencing by in ovo electroporation.* Methods Mol Biol, 2008. **442**: p. 231-44.
516. Baeriswyl, T. and E.T. Stoeckli, *In ovo RNAi opens new possibilities for temporal and spatial control of gene silencing during development of the vertebrate nervous system.* J RNAi Gene Silencing, 2006. **2**(1): p. 126-35.
517. Bregenhorn, S., et al., *Non-canonical uracil processing in DNA gives rise to double-strand breaks and deletions: relevance to class switch recombination.* Nucleic Acids Res, 2016. **44**(6): p. 2691-705.
518. Luby, T.M., et al., *The mu switch region tandem repeats are important, but not required, for antibody class switch recombination.* The Journal of experimental medicine, 2001. **193**(2): p. 159-68.
519. Min, I.M., et al., *The Smu tandem repeat region is critical for Ig isotype switching in the absence of Msh2.* Immunity, 2003. **19**(4): p. 515-24.
520. Wu, X. and J. Stavnezer, *DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination.* The Journal of experimental medicine, 2007. **204**(7): p. 1677-89.
521. Phan, R.T. and R. Dalla-Favera, *The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells.* Nature, 2004. **432**(7017): p. 635-9.
522. Kuppers, R., *Mechanisms of B-cell lymphoma pathogenesis.* Nature reviews. Cancer, 2005. **5**(4): p. 251-62.
523. Jankovic, M., A. Nussenzweig, and M.C. Nussenzweig, *Antigen receptor diversification and chromosome translocations.* Nature immunology, 2007. **8**(8): p. 801-8.
524. Baron, B.W., et al., *Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(11): p. 5262-6.
525. Iida, S., et al., *The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene.* Blood, 1996. **88**(11): p. 4110-7.
526. Neri, A., et al., *B cell lymphoma-associated chromosomal translocation involves candidate oncogene lyt-10, homologous to NF-kappa B p50.* Cell, 1991. **67**(6): p. 1075-87.
527. Ye, B.H., et al., *Cloning of bcl-6, the locus involved in chromosome translocations affecting band 3q27 in B-cell lymphoma.* Cancer research, 1993. **53**(12): p. 2732-5.
528. Nussenzweig, A. and M.C. Nussenzweig, *Origin of chromosomal translocations in lymphoid cancer.* Cell, 2010. **141**(1): p. 27-38.
529. Kuppers, R. and R. Dalla-Favera, *Mechanisms of chromosomal translocations in B cell lymphomas.* Oncogene, 2001. **20**(40): p. 5580-94.
530. Ruiz, J.F., B. Gomez-Gonzalez, and A. Aguilera, *AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants.* PLoS genetics, 2011. **7**(2): p. e1002009.

531. Difilippantonio, M.J., et al., *DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation.* Nature, 2000. **404**(6777): p. 510-4.
532. Gao, Y., et al., *Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development.* Nature, 2000. **404**(6780): p. 897-900.
533. Jones, S., *An overview of the basic helix-loop-helix proteins.* Genome Biol, 2004. **5**(6): p. 226.
534. Fairman, R., et al., *Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10429-33.
535. Prendergast, G.C., D. Lawe, and E.B. Ziff, *Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation.* Cell, 1991. **65**(3): p. 395-407.
536. Prendergast, G.C. and E.B. Ziff, *Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region.* Science, 1991. **251**(4990): p. 186-9.
537. Xu, L., et al., *Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation.* Angew Chem Int Ed Engl, 2014. **53**(42): p. 11223-7.
538. Crawford, D.J., et al., *Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism.* J Am Chem Soc, 2016. **138**(3): p. 730-3.
539. Tamanaha, E., et al., *Distributive Processing by the Iron(II)/alpha-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases.* J Am Chem Soc, 2016. **138**(30): p. 9345-8.
540. Munzel, M., et al., *Quantification of the sixth DNA base hydroxymethylcytosine in the brain.* Angew Chem Int Ed Engl, 2010. **49**(31): p. 5375-7.
541. Pfaffeneder, T., et al., *The discovery of 5-formylcytosine in embryonic stem cell DNA.* Angew Chem Int Ed Engl, 2011. **50**(31): p. 7008-12.
542. Raiber, E.A., et al., *Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase.* Genome Biol, 2012. **13**(8): p. R69.
543. Quivoron, C., et al., *TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis.* Cancer Cell, 2011. **20**(1): p. 25-38.
544. Figueroa, M.E., et al., *Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.* Cancer Cell, 2010. **18**(6): p. 553-67.

ACKNOWLEDGEMENTS

First, I want to thank **Joe Jiricny** for giving me the chance to do my PhD in his lab and for being an extremely supporting and motivating boss. His passion for science and never ending curiosity was highly contagious during my whole PhD. WIth great devotion, he made his lab and the whole IMCR an inspiring place to work with an exceptional atmosphere of helpfulness and care.

Thank you also to my committee members **Ueli**, **Primo** and **Jörn** for helpful discussions and advice during the meetings and for always being very encouraging.

Many thanks to **Andrea**, **Weihong** and **Giancarlo** from the FGCZ for their continuous support with the sequencing and data analysis.

A warm thank you goes to my fellow sufferers **Julia**, **Tine**, **Martin**, **Cosimo** and **Maryna**, who made this time so memorable, in the lab, as well as outside of the lab. They were always there for mental support or some distraction.

Then, I want to thank the present members of the lab at ETH; **Sara**, **Saho**, **Shun** and most of all **Rachel**, who is a invaluable help in the methylation project and manages the lab with great commitment.

Many thanks also to the past members of the Jiricny group; **Maite, Steffi, Anja**, **Svenja**, **Simone**, **Mariela** and **Antonio**, for being good friends and a great support in the lab.

I am very grateful to the people at the IMCR for being very helpful and friendly and always open for little chats in the hallway. Special thanks go to **Farah** and **Odete** for caring so much for the institute and making our lives so easy.

My biggest thank goes to my family and friends. They support me in whatever I do, without knowing what I actually do. They were always there to distract me when things were not going well in the lab and reminded me of the important things in life.

CURRICULUM VITAE

LIA KALLENBERGER

Steinstrasse 23
CH-8003, Zurich
+41 79 258 82 89
lia@kallenberger.ch

Date of birth	23.08.1989
Nationality	Swiss, Amriswil TG
Marital status	unmarried

EDUCATION

University of Zurich 01/2014 - 05/2017
PhD in cancer biology
Fasttrack program of the University of Zurich
Cancer Biology PhD Program of the Life Science Zurich Graduate School (UZH & ETHZ)

Thesis: "Interplay of DNA metabolic pathways involved in the regulation of gene expression and genomic rearrangements"

University of Zurich 08/2012 - 12/2013
Master in Cell -and Molecular Biology
Fasttrack Program of the University of Zurich

Thesis: "Elucidating the role of mismatch repair in the induction of DNA double strand breaks during class switch recombination"

University of Zurich 09/2009 - 07/2012
Bachelor in Biology

Kantonsschule Wiedikon 08/2002 - 08/2008
Matura, Modern languages

RELEVANT WORK EXPERIENCE

Zweifel Pomy-Chips AG 04/2009 – 12/2013
Internship in quality control
- Hourly quality control of the chips in production
- Precise measurements and analysis of the parameters

University Hospital Zurich 10/2008 –03/2009
Internship in the molecular oncology department
- Investigation of biomarkers for the differentiation of mesothelioma and Non-small-cell lung cancer

University Hospital Bern 09/2008
Internship in the research lab of the intensive care unit
- Introduction into cell culture work and molecular biology techniques

CERTIFICATES

GCP	**Good Clinical Practice course** Clinical Trials Center CTC Center for Clinical Research ZKF	2015
LTK Module 20	**Animal course** Introductory Course in Laboratory Animal Science Aquatic animals Vetsuisse-faculty	2014
CAE	**University of Cambridge** Certificate in Advanced English	2008

SELECTED CONFERENCES

Swiss Meeting on Genome Stability & Chromatin Dynamics, Switzerland 05/2014
Poster presentation: "Investigation of the molecular mechanisms of 5-methylcytosine metabolism"

Student Retreat, Cancer PhD Program, Switzerland 02/2014
Poster presentation: "Investigation of the molecular mechanisms of 5-methylcytosine metabolism"

Gordon Conference on Mutagensis, Girona 06/2016
Poster presentation: "Unravelling the mechanism of AID-induced translocations by analysis of the interplay between mismatch repair and base excision repair"

Epigenetics and Chromatin, Cold Spring Harbor 09/2016
Poster presentation: "Investigation of the role of TET-induced cytosine modification on transcription of an estrogen-responsive gene"

PUBLICATIONS

"Mismatch repair-dependent metabolism of 06-methylguanine-containing DNA in Xenopus laevis egg extracts." *Olivera Harris M, Kallenberger L, Artola Boran M. Enoiu M, Constanzo V, Jiricny J.* DNA Repair 2015

"Non-canonical uracil processing in DNA gives rise to double-strand breaks and deletions: relevance to class switch recombination." *Bregenhorn S, Kallenberger L, Artola Boran M, Peña-Diaz J, Jiricny J.* NAR 2016

"A methylation-responsive E-box binding protein controls estrogen-dependent transcription of the chicken vitellogenin gene." *Kallenberger L, Erb R, Jiricny J.* Manuscript in preparation.

LANGUAGES

German: mother tongue
English: very good knowledge, oral and written
French: good knowledge, oral and written
Spanish: good knowledge, oral and written

PERSONAL INTERESTS

Climbing, snowboard, hiking, traveling, reading, cooking

APPENDIX

Mismatch repair-dependent metabolism of O^6-methylguanine-containing DNA in *Xenopus laevis* egg extracts

Maite Olivera Harris[a,b], Lia Kallenberger[a], Mariela Artola Borán[a], Milica Enoiu[a], Vincenzo Costanzo[c], Josef Jiricny[a,b,*]

[a] Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[b] Department of Biology, Swiss Federal Institute of Technology (ETH), Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[c] IFOM-European Institute of Oncology Campus, Via Adamello 16, 20139 Milano, Italy

I contributed to this work by performing MMR assays and Western Blots.

DNA Repair 28 (2015) 1–7

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

Brief Communication

Mismatch repair-dependent metabolism of O^6-methylguanine-containing DNA in *Xenopus laevis* egg extracts

Maite Olivera Harris[a,b], Lia Kallenberger[a], Mariela Artola Borán[a], Milica Enoiu[a], Vincenzo Costanzo[c], Josef Jiricny[a,b,*]

[a] Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[b] Department of Biology, Swiss Federal Institute of Technology (ETH), Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[c] IFOM-European Institute of Oncology Campus, Via Adamello 16, 20139 Milano, Italy

ARTICLE INFO

Article history:
Received 29 April 2014
Received in revised form 15 January 2015
Accepted 20 January 2015
Available online 4 February 2015

Keywords:
DNA damage signaling
DNA replication
O^6-methylguanine
Mismatch repair
S_N1-type alkylating agents
Xenopus laevis egg extracts

ABSTRACT

The cytotoxicity of S_N1-type alkylating agents such as *N*-methyl-*N*′-nitrosourea (MNU), *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG), or the cancer chemotherapeutics temozolomide, dacarbazine and streptozotocin has been ascribed to the persistence of O^6-methylguanine (^{me}G) in genomic DNA. One hypothesis posits that ^{me}G toxicity is caused by futile attempts of the mismatch repair (MMR) system to process ^{me}G/C or ^{me}G/T mispairs arising during replication, while an alternative proposal suggests that the latter lesions activate DNA damage signaling, cell cycle arrest and apoptosis directly. Attempts to elucidate the molecular mechanism of ^{me}G-induced cell killing *in vivo* have been hampered by the fact that the above reagents induce several types of modifications in genomic DNA, which are processed by different repair pathways. In contrast, defined substrates studied *in vitro* did not undergo replication. We set out to re-examine this phenomenon in replication-competent *Xenopus laevis* egg extracts, using either phagemid substrates containing a single ^{me}G residue, or methylated sperm chromatin. Our findings provide further support for the futile cycling hypothesis.

1. Introduction

The postreplicative mismatch repair (MMR) system improves replication fidelity by up to three orders of magnitude through excising nucleotides misincorporated into the nascent strand by DNA polymerases [1,2]. This task is carried out predominantly by the mismatch recognition protein MutSα (a heterodimer of MSH2 and MSH6), the endonuclease MutLα (MLH1/PMS2 heterodimer) and EXO1 exonuclease, together with accessory replication factors PCNA, RPA, RFC, polymerase-δ and DNA ligase I [3,4].

In contrast to its key role as an antimutator [5], MMR can perturb genomic stability by attempting to process DNA damage induced by, for example, S_N1-type methylating agents such as *N*-methyl-*N*′-nitrosourea (MNU), *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG) or temozolomide [6,7]. The toxicity of this class of chemicals, some of which (e.g. temozolomide, dacarbazine, streptozotocin) are used in the therapy of various types of cancer, is linked to their ability to generate O^6-methylguanine (^{me}G) in DNA. This simple modification can be efficiently detoxified by O^6-methylguanine DNA-methyltransferase (MGMT), which transfers the methyl group from the guanine onto a cysteine residue in its active site [8]. Because MGMT is inactivated by this transaction, and because some cells express only low amounts of this protein, ^{me}G can persist in genomic DNA after treatment with these chemicals and cause cytotoxicity. More than three decades ago, Karran and Marinus showed that ^{me}G toxicity in *E. coli* was dependent on functional MMR [9]. Like MMR-deficient bacteria, mammalian cells lacking MMR are also highly resistant to killing by S_N1-type methylating agents [7,10–12], but the molecular events leading to cell death have been the subject of extensive discussion. One hypothesis, referred to as "futile cycling" [7], posits that replicating polymerases will incorporate a C or a T opposite the ^{me}G. As ^{me}G/C and ^{me}G/T are not Watson-Crick base pairs, they will activate MMR. However, because MMR is directed to the nascent strand, EXO1 will catalyze degradation of the strand containing the mispaired C or T. As the ^{me}G residue will remain in the template strand, repair synthesis will regenerate the ^{me}G/C or ^{me}G/T mispairs that will trigger further, "futile" rounds of MMR-directed excision and

* Corresponding author at: Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.
Tel.: +41 044 6353450.
E-mail address: jiricny@imcr.uzh.ch (J. Jiricny).

http://dx.doi.org/10.1016/j.dnarep.2015.01.014
1568-7864/

resynthesis, which were suggested to activate cell cycle arrest and apoptosis. Support for this hypothesis comes from several directions; iterative cycles of MMR-dependent excision and repair synthesis were seen in *in vitro* assays [13], and cell cycle arrest induced by treatment with S_N1-type methylating agents was shown to require two rounds of replication [14–16]. In an attempt to explain the latter finding, we suggested that the MMR machinery might abandon its futile attempts at repair of the ^{me}G/C or ^{me}G/T mispairs and that DNA synthesis might re-initiate downstream. This would leave single-stranded gaps opposite ^{me}G, the collision of which with replication forks in the subsequent S phase would give rise to double-strand breaks that would activate the ATR and ATM checkpoint machineries and cause a cell cycle arrest in the second G_2/M [16]. In our later work, we succeeded in identifying persistent gaps in nascent DNA of cells treated with low MNNG doses [14], which provided further support for the above pathway.

Although the "futile cycling" hypothesis [7] has met with broad acceptance, work from the Hsieh laboratory suggested that MutSα may act as a damage sensor that recruits and activates ATR directly [17], and support for this hypothesis was provided by evidence of interactions between MMR proteins and the ATR checkpoint machinery [18,19]. Although this "direct signaling" model fails to explain why recognition of the ^{me}G-containing mismatches should activate cell cycle arrest only in the second S phase, there is no reason why the two scenarios should not co-exist, each addressing a subset of lesions.

The experimental evidence described above has been obtained either from *in vitro* MMR assays [13,17], or from cell culture experiments [14–16]. The former system deploys heteroduplex substrates that do not undergo replication in the human cell extracts, while the latter is not amenable to manipulation. We therefore decided to analyze the processing of ^{me}G-containing substrates in *Xenopus laevis* egg extracts (XEE), an experimental system that permits the study of DNA damage signaling, DNA replication and DNA repair either in plasmid substrates or in chromatin, under conditions that can be altered at will [20,21]. We show here that replication of ^{me}G-containing DNA in XEE did indeed give rise to persistent single-stranded gaps opposite ^{me}G. Furthermore, recruitment of MMR proteins to the ^{me}G-containing DNA occurred throughout replication and persisted thereafter, which is likely to be indicative of repeated attempts to repair the gaps. Importantly, recruitment of MMR proteins to alkylation-damaged chromatin did not coincide with that of the ATR checkpoint machinery, and we detected no evidence of checkpoint activation after the single replication round. Our data thus provide further support for the "futile cycling" hypothesis.

2. Materials and methods

Detailed information regarding antibodies, primers, DNA substrates, *X. laevis* egg extract preparation, chromatin isolation and MSH6 immunodepletion, as well as band-shift-, dot blot- and MMR assays in *X. laevis* extracts (xMMR) can be found in Supplementary Information.

2.1. X. laevis egg extracts and chromatin isolation

Lysolecithin-permeabilised *X. laevis* sperm nuclei ("sperm chromatin") were prepared as described [22]. S phase extract capable of performing a single round of replication was prepared as previously described [23]. Briefly, eggs were dejellied, activated with calcium ionophore (Sigma), rinsed with S buffer (50 mM Hepes-KOH pH 7.5, 50 mM KCl, 2.5 mM $MgCl_2$ and 250 mM sucrose), transferred to 2 ml Eppendorf tubes and crushed by centrifugation for 12 min at 13′200 rpm. The cytoplasmic layer was removed and, after addition of CytB (Sigma), cleared by centrifugation for 25 min at 70′000 rpm (Beckman TL199 ultracentrifuge, TL55 swing-bucket rotor). The extract was supplemented with 250 μg/μl cycloheximide, 25 mM phosphocreatine and 10 μg/ml creatine phosphokinase before use.

Where indicated, sperm chromatin was exposed to varying doses of MNU, MNNG or 1% MMS (all from Sigma) for 30 min at RT.

HSS and NPE were prepared as described previously [24]. To make NPE, unfractionated egg extract (crude S phase extract) was supplemented with nocodazole (3.3 μg/ml), cycloheximide (50 μg/ml), DTT (1 mM), aprotinin/leupeptin (10 μg/ml each) and cytochalasin B (5 μg/ml), and centrifuged at 10′000 rpm for 15 min in a Sorvall HB-6 rotor at 4 °C. The dark brown layer at the top of the tube was removed by aspiration and the extract was decanted into a fresh Falcon tube, supplemented with an ATP regeneration system and demembranated sperm chromatin to a concentration of 4000 n/μl. The mixture was incubated at room temperature for 75–90 min. To collect the nuclei, the reaction was centrifuged for 2 min at 10,000 rpm at 4 °C in a Sorvall HB-6 rotor. The clear ~4 mm layer of nuclei was removed from the top of the tube and transferred to 5 × 20 mm ultracentrifuge tubes. It was then centrifuged for 30 min at 55,000 rpm in a Beckmann TL55 rotor. After centrifugation, lipids at the top of the sample were aspirated and the clear nucleoplasm was harvested.

2.2. Isolation of nuclear and chromatin fractions

To isolate sperm chromatin, sperm nuclei (4000 n/μl) were added to S phase extract for the indicated times. For Western blotting, samples were diluted in ten volumes of EB buffer (100 mM KCl, 2.5 mM $MgCl_2$ and 50 mM Hepes-KOH pH 7.5) containing 0.25% Nonidet P-40 and centrifuged through a 30% sucrose layer at 10,000 rpm for 5 min at 4 °C using a Sorvall HB-6 rotor, washed twice with 500 μl EB buffer and centrifuged at 10′000 rpm for 1 min. The pellet was resuspended in Laemmli loading buffer, the proteins were resolved by either 4–15%, 7.5% or 10% SDS-PAGE and analyzed by Western blotting with specific antibodies as indicated. Unless otherwise indicated, Precision plus protein dual color standard (Biorad) was used as a size reference. Nuclei were isolated by the same procedure, except that the EB buffer contained no detergent.

2.3. DNA replication

Replication of sperm chromatin in S phase extract (4000 n/μl) was measured by monitoring the incorporation of [^{32}P]dCMP for 120 min at 23 °C as previously described [25]. Treatments are indicated in the figure legends. 5 μl aliquots were removed and combined with 15 μl of STOP mix (6 mM EDTA, 0.13% phosphoric acid, 10% Ficoll, 5% SDS, 0.2% bromophenol blue and 80 mM Tris–HCl pH 8). Proteinase K (Roche) was then added to a final concentration of 2 μg/μl for 1 h at 37 °C. The samples were separated on 0.8% agarose/TAE gels, and the dried gels were exposed to PhosphoImager screens. The images were quantitated with ImageQuant software (GE Healthcare).

For plasmid replication, plasmid DNA pRichi2850t was initially incubated in 6 μl HSS for 30 min at 23 °C, followed by addition of a 2-fold volume of NPE. DNA replication was assayed by [α-^{32}P]dCMP (^{32}P) incorporation as described [24]. For substrate recovery, 40 μl STOP solution (5% SDS, 80 mM Tris–HCl pH 8 and 5 mg/ml RNase) were added for 30 min at 37 °C. The reaction was then incubated with Proteinase K (1 mg/ml) overnight at 37 °C. Samples were purified on MiniElute columns (Qiagen) and subjected to restriction digests in the presence of RNase (0.5 mg/ml, 5prime). The digested DNA was cleaned up again and resolved on 1% TAE agarose gels stained with GelRed.

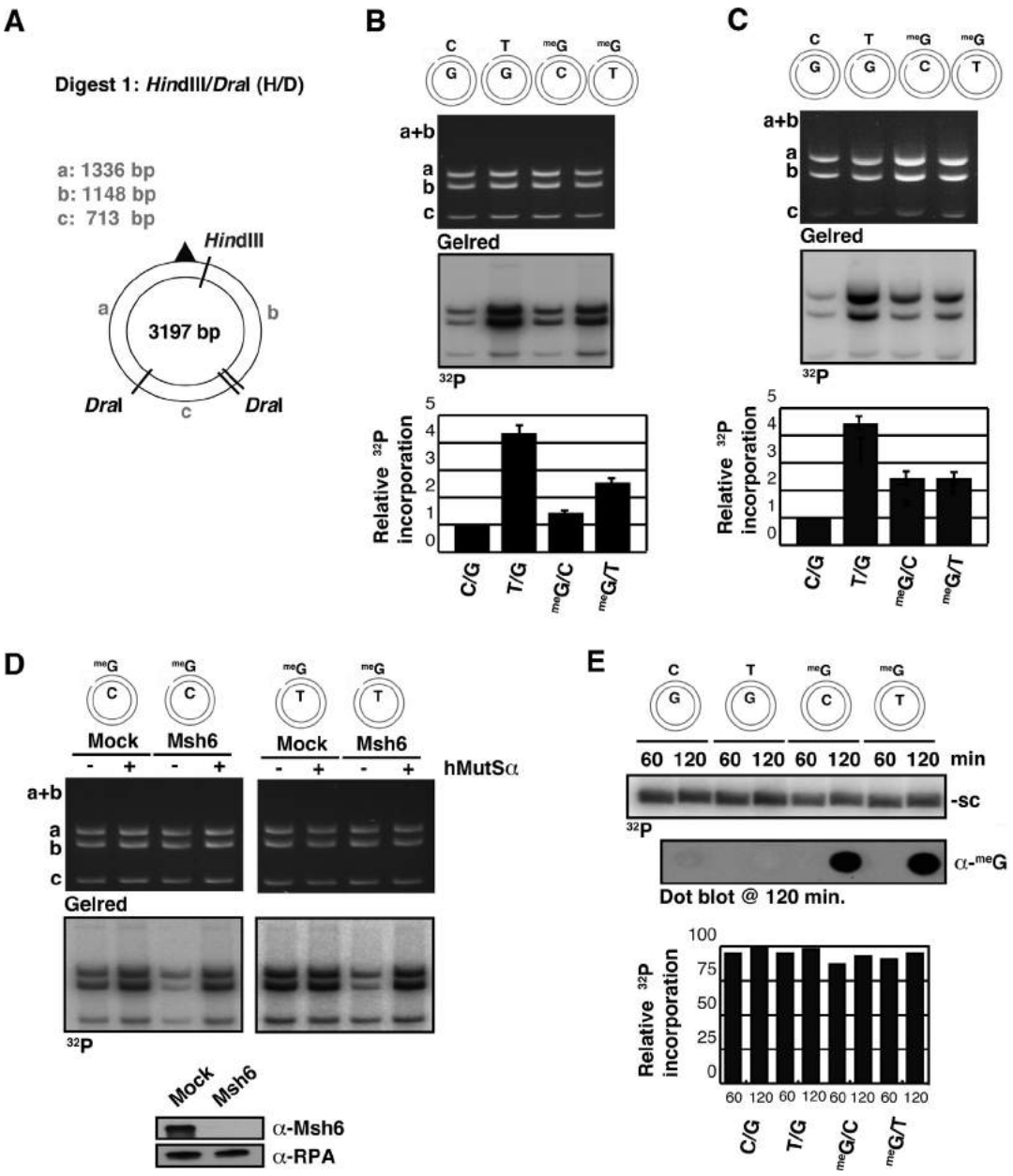


Fig. 1. Replication-dependent and -independent MMR activity in *X. laevis* egg extracts. (A) Schematic representation of the phagemid substrate and the fragments generated by *Hin*dIII/*Dra*I digestion. The *Hin*dIII site lies 12 bp from the *Sal*I site that contains the mismatch (triangle). Nt.*Bst*NBI generates a single nick in the 3197 bp substrate (not depicted), either 363 nucleotides 5′ from the C, T or ^{me}G residue in the top (complementary) strand, or 376 nucleotides 3′ from the mispaired G, C or T in the bottom (viral) strand. The MMR excision/resynthesis machinery is directed to the nicked strand. (B and C) MMR assays in Patrin-2-pretreated HSS extracts with top- (B) or bottom- (C) nicked C/G, T/G, ^{me}G/C and ^{me}G/T substrates. The phagemids recovered from the reaction were digested with *Hin*dIII/*Dra*I, and the restriction fragments were separated on 1% agarose/TAE gels stained with GelRed. The dried gels were then autoradiographed (^{32}P). Bar graphs represent data from three independent experiments, with error bars showing standard deviation from the mean. (D) As in (B), but the ^{me}G/C or ^{me}G/T substrate was incubated with Patrin-2-pretreated HSS extract, either mock- or Msh6-depleted. The MMR defect in the Msh6-depleted extracts could be rescued by the addition of recombinant human MutSα. Bottom: Western blot showing Msh6 depletion efficiency in the extract. (E) Kinetic analysis of phagemid replication (supercoiled product, sc) in Patrin-2-pretreated *X. laevis* extract. The indicated substrates were incubated in HSS for 30 min to allow the formation of a pre-replication complex, followed by addition of NPE that initiates replication (see Supplemental Information for details). Aliquots were withdrawn at the indicated time points, cleaned up and loaded on 0.8% agarose/TAE gels. The dried gels were then autoradiographed (^{32}P). Middle panel: dot blot analysis of the replicated substrates with an anti-^{me}G antibody. Bottom panel: bar graph depicting data quantification from the top panel.

All replication reactions were carried out in the presence of Patrin-2 [O^6-(4-bromoethynyl)guanine; 100–150 μM], unless otherwise indicated.

3. Results and discussion

We first set out to examine MMR activity in *X. laevis* egg extracts, using circular heteroduplex substrates (Fig. 1A) containing either a single T/G, ^{me}G/C or ^{me}G/T mismatch. The substrates, as well as the control C/G homoduplex, were generated by primer extension on single-stranded phagemid DNA, using synthetic oligonucleotides as primers. The mispairs were situated within the unique *Sal*I restriction site (Fig. S1A), which makes the DNA refractory to cleavage by this enzyme.

Chemical synthesis of oligonucleotides requires the presence of an isopropyl protecting group on the N^2 position of guanines, which

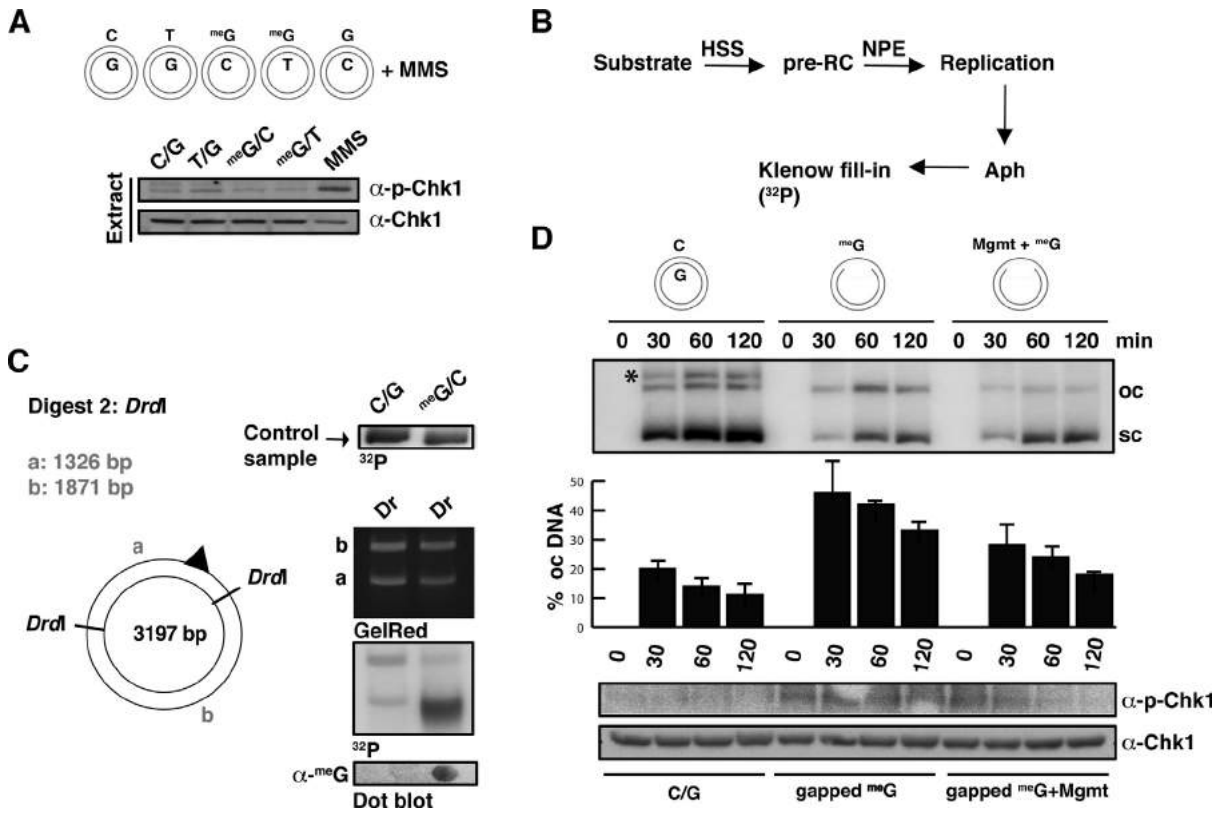


Fig. 2. Replication of ^{me}G-containing DNA in *X. laevis* egg extracts gives rise to gaps that escape checkpoint monitoring but hinder further replication. (A) Analysis of checkpoint activation after replication of the indicated substrates in Patrin-2-pretreated HSS + NPE. Aliquots removed at the 120 min time point were resolved by 7.5% SDS-PAGE and probed for phospho-Chk1 (S_{345}, p-Chk1) and total Chk1. (B) Scheme of post-replicative gap-filling assays. (C) Schematic representation of the substrate digested with *Drd*I (left panel) and gap-filling reactions of the C/G and ^{me}G/C substrates (top right panel). Substrates were replicated in Patrin-2-pretreated extract without addition of radioactive nucleotide (control sample was spiked with [α-^{32}P]dCTP). Aphidicolin (7.5 μM) was added for 5 min after the replication was complete, the substrate was isolated and the gap-filling reaction with Klenow polymerase was performed in the presence of [α-^{32}P]dCTP. The substrates were then digested with *Drd*I (Dr) and the samples were separated on 1% agarose/TAE gels stained with GelRed. The dried gels were subsequently autoradiographed (^{32}P). Bottom panel: dot blot of the replicated substrates with an anti-^{me}G antibody. (D) Replication of the gapped ^{me}G/C and the control closed-circular C/G substrates in HSS + NPE supplemented with [α-^{32}P]dCTP. Aliquots removed at the indicated times were divided in two, one portion was used to extract the DNA, the other was used for protein analysis. Top panel: autoradiograph of DNA molecules recovered from the replication reactions and separated on 0.8% agarose/TAE gels. oc – open circular, sc – supercoiled. Asterisk denotes supercoiled dimer C/G molecules. These are not visible in the gapped substrate reactions due to the substantially lower yield of supercoiled plasmids. Bar graph represents data from three independent experiments, with error bars showing standard deviation from the mean. Bottom panel: Western blot of extract proteins resolved by 7.5% SDS-PAGE and probed for phoshpo-Chk1 (p-Chk1) and total Chk1.

is removed in the final deprotection step by ammonia treatment. The latter step cannot be used on oligonucleotides containing O^6-methylguanine, because ammonia would remove the 6-methoxy group. An alternative deprotection is therefore used, but this is often incomplete. In order to ensure that our oligonucleotides indeed contained fully deprotected ^{me}G, we annealed them to their complementary sequences to generate 31-mer duplexes containing ^{me}G/C or ^{me}G/T mispairs, and carried out gel-shift experiments with purified human MutSα. Neither oligonucleotide was bound efficiently, but, upon treatment with recombinant *X. laevis* Mgmt (Fig. S1B), a strong gel-shift was observed with the ^{me}G/T oligonucleotide, which is indicative of its efficient conversion to G/T (Fig. S1C).

We next had to inhibit the Mgmt activity in the extracts. Human MGMT is routinely inhibited by O^6-benzylguanine (BG), and this was the case also for the *X. laevis* enzyme. As shown in Fig S1D (left panel), the ^{me}G/C substrate incubated with the high speed supernatant (HSS) in the presence of 10 μM BG was readily converted to G/C, as shown by *Sal*I cleavage of the plasmid into fragments **a** and **b**. Higher concentrations inhibited the cleavage, which showed that the ^{me}G/C mismatch was intact. However, the high BG concentrations necessary to inhibit Mgmt in the extracts inhibited also the MMR efficiency (data not shown). For this reason, we made use of Patrin-2 [O^6-(4-bromoethynyl)guanine; a kind gift of Geoff Margison], which inhibits Mgmt with similar efficiency to BG (Fig. S1D, right panel), but without the deleterious effect on MMR.

In earlier studies, ^{me}G in DNA was shown to be addressed by MMR in *E. coli* [26], *S. cerevisiae* [27] and human cells [13,28], and we set out to extend these findings to the *X. laevis* system. Although the ^{me}G/C or ^{me}G/T oligonucleotide duplexes were not efficiently bound by purified MutSα in the extracts (Fig. S1C), the circular phagemids containing a ^{me}G/C or ^{me}G/T mispair and a single nick in either the complementary (Fig. 1B) or the viral strand (Fig. 1C) activated excision and repair synthesis, as shown by incorporation of [^{32}P]dCMP predominantly into fragments **a** and **b** of the phagemid DNA. This process was dependent on MutSα, as shown by immunodepletion of this factor from the extracts with a *X. laevis* anti-Msh6 antibody (Fig. S1E–G) and subsequent rescue with purified recombinant human MutSα (Fig. 1D).

We next set out to ask whether the presence of ^{me}G in DNA affects replication. As shown in Fig. 1E, the covalently closed T/G, ^{me}G/C, ^{me}G/T and the C/G control phagemids were replicated efficiently in a mixture of high speed supernatant (HSS) and nucleoplasmic extracts (NPE) of *X. laevis* eggs, with the ^{me}G-containing substrates showing only a slight decrease in [^{32}P]dCMP incorporation (5–13%). Importantly, ^{me}G could be shown to persist in the ^{me}G/C and ^{me}G/T substrates after replication was completed

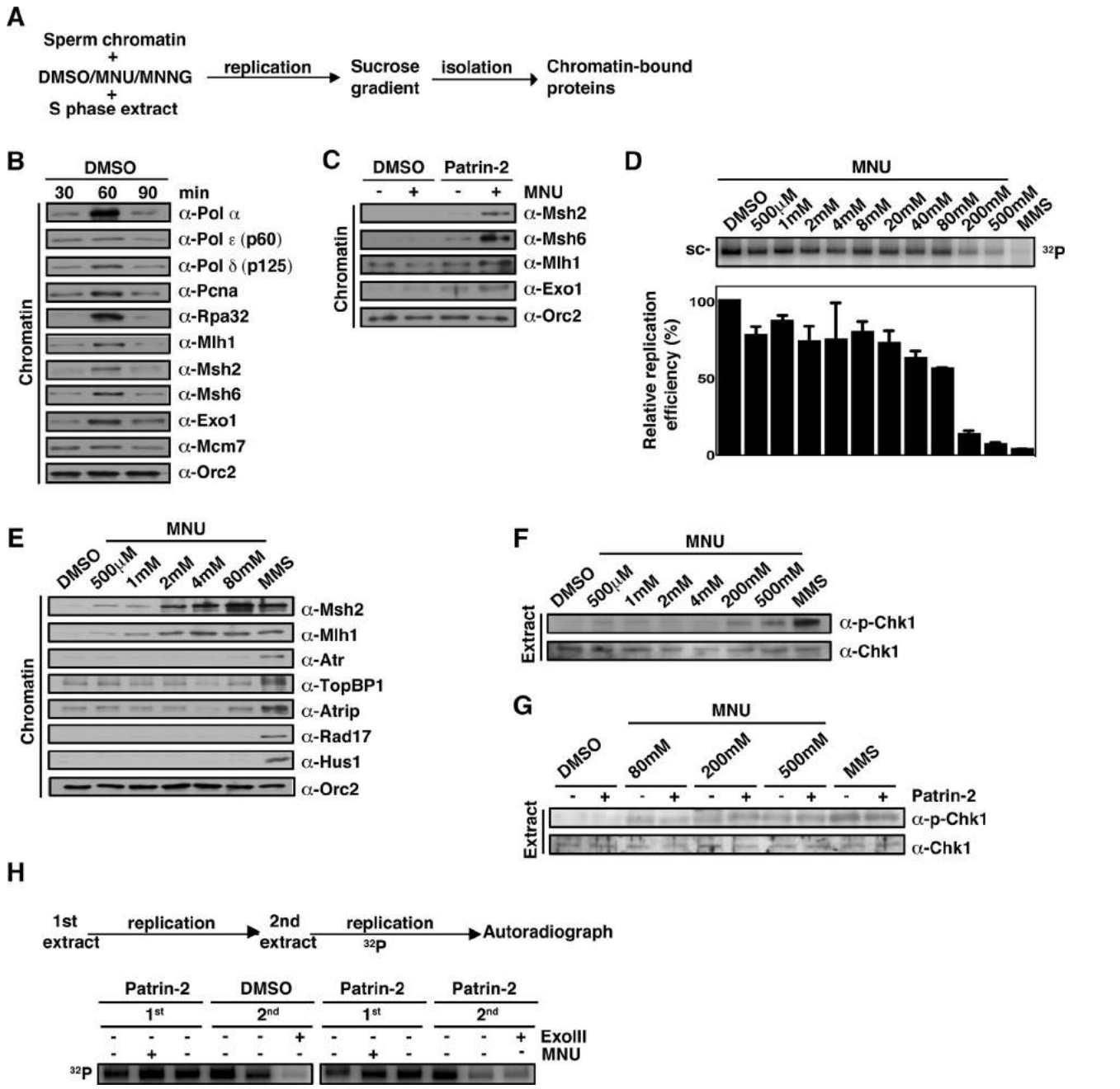


Fig. 3. Recruitment of *X. laevis* MMR proteins to MNU-treated sperm chromatin. (A) Schematic representation of the chromatin association assay. (B) Western blot of recruitment of replication- and MMR proteins to sperm chromatin treated with 1.5 mM MNU. (C) Western blot of recruitment of MMR proteins to undamaged or MNU-treated (1.5 mM) sperm chromatin, incubated for 90 min with extracts pretreated with DMSO (control) or Patrin-2 (150 μM). (D) Efficiency of replication of MNU-treated sperm chromatin in S phase extracts, monitored by supplementing the extracts with [α-^{32}P]dCTP (^{32}P) and incubation for 120 min (sc-supercoiled DNA). Top panel: autoradiograph of a 0.8% agarose/TAE gel (^{32}P). Bottom panel: column graph of data from three independent experiments, with error bars representing standard deviation from the mean. (E) Western blot showing recruitment of MMR- and checkpoint proteins to sperm chromatin treated with increasing concentrations of MNU. Extracts were pretreated with Patrin-2. MMS (1%) treatment was used as a positive control of checkpoint activation and Orc2 was used as a loading control. (F) Phosphorylation of Chk1 (S_{345}) in Patrin-2-pretreated S phase extracts in response to MNU. Nuclei were isolated after 120 min and probed for p-Chk1 and total Chk1. (G) MMR-independent phosphorylation of Chk1 in S phase extracts in response to high MNU concentrations. Nuclei were isolated after 120 min and probed for p-Chk1 and total Chk1. (H) Re-replication of MNU-treated sperm chromatin. As shown in the scheme, the chromatin was first replicated in Patrin-2 pretreated extracts, isolated, and incubated with fresh extracts that were either pretreated with Patrin-2 or only with DMSO. The extracted DNA was separated on 0.8% agarose/TAE gels and autoradiographed (^{32}P). ExoIII-pre-treated chromatin was used as positive control; this treatment generates single-stranded gaps from random strand breaks in the DNA.

(120 min), as confirmed in the dot blot assay using a specific antibody against ^{me}G (a kind gift of Jürgen Thomale).

To test whether replication of ^{me}G-containing substrates induced DNA damage signaling in the extracts, we incubated the heteroduplexes with HSS and NPE as above. The control, C/G phagemid, either untreated or treated with MMS (commonly used to trigger DNA damage signaling in *X. laevis* extracts), was used as a control. As shown in Fig. 2A, only the MMS-treated C/G substrate gave rise to significant ATR activation upon replication completion (120 min), as seen by phosphorylation of the ATR downstream target Chk1. Closer analysis of this signaling indicated that replication of the ^{me}G/C substrate also activated ATR at the 30 min. time point, but the phospho-Chk1 signal did not persist until replication was completed (Fig. S2A). It is possible that this transient activation

reflected an interaction of MMR with the checkpoint machinery as observed previously both *in vitro* [17,18] and *in vivo* [19,29].

As mentioned above, DNA damage signaling induced by S_N1 type methylating agents in human cells is activated after the first S phase, but cell cycle arrest occurs in the second G_2/M phase [10,16], believed to be triggered by single-stranded gaps resulting from replication re-initiation behind the unrepaired ^{me}G residues [14]. We therefore searched for the presence of such gaps in our *in vitro*-replicated substrates. We designed an experimental protocol shown schematically in Fig. 2B, in which we incubated the C/G and ^{me}G/C phagemids with HSS and NPE, isolated the replicated DNA, incubated it with Klenow polymerase in the presence of aphidicolin, dATP, dGTP, dTTP and [α-^{32}P]dCTP, and then digested it with *Drd*I. As shown in Fig. 2C, only the ^{me}G/C substrate incorporated substantial amounts of [^{32}P]dCMP into fragment **a**. Because this fragment contains the ^{me}G residue, it is most likely that the radionucleotide was incorporated into gaps remaining opposite this modified base after replication. This finding was complemented by the fact that the non-denatured phagemid isolated from the replication mixture reacted with an anti-^{me}G antibody in a dot blot assay (Fig. 2C, lowest panel).

The futile cycling hypothesis [7] posits that the single-stranded gaps opposite ^{me}G would be effectively irreparable, because polymerases attempting gap-filling would be aborted by MMR activation [13]. We therefore set out to test this prediction in *X. laevis* egg extracts. We anticipated that replication of the control C/G plasmid would give rise predominantly to supercoiled molecules, whereas replication of the ^{me}G/C plasmid in which a gap had been generated opposite the methylated guanine (Fig. S2B–D) would give rise to open-circular molecules that should be converted to supercoiled DNA with only low efficiency. However, pretreatment of the latter gapped plasmid with Mgmt, which will have converted ^{me}G to G, should give rise to proportionally more supercoiled plasmid DNA. This was indeed the case, as shown in Fig. 2D. Interestingly, the gapped ^{me}G/C phagemid did give rise to supercoiled molecules at the later time points, which indicates that the gaps were eventually repaired. This would happen if a DNA polymerase succeeded in filling the gap completely such that the 3′ terminus of the resynthesized strand could be ligated to the 5′ terminus of the downstream nascent DNA strand generated by replication reinitiation. In this scenario, the ^{me}G would still be in the DNA, but the duplex would be largely refractory to MMR, as there would be no strand discontinuity in the vicinity. We also detected ATR activation, as shown by phosphorylation of Chk1 that appeared already prior to replication, as seen previously with gapped substrates [30,31]. Importantly, in the case of the gapped ^{me}G plasmid, the signal persisted throughout the 120 min time course of the assay, but not when the same plasmid was pre-treated with Mgmt (Fig. 2D, bottom panel).

In order to gain more information about the interplay between replication and MMR-dependent DNA damage signaling in a more physiologically relevant setting, we studied the recruitment of the respective proteins to *X. laevis* sperm chromatin, as shown schematically in Fig. 3A. Using in-house-generated rabbit polyclonal antibodies against frog polypeptides (Fig. S1E–G), as well as other antibodies generously provided by others (see M&M), we first monitored protein recruitment to undamaged chromatin in S phase extracts. As shown in Fig. 3B, Msh2, Msh6, Mlh1 and Exo1 were recruited with kinetics similar to those of the three replicative polymerases and several other replication factors, which confirms that MMR and replication are closely associated [32,33]. When we repeated the experiment with chromatin treated with 1.5 mM MNU (Fig. 3A and C), MMR protein recruitment was significantly increased when the chromatin was incubated with extracts pre-treated with the Mgmt inhibitor Patrin-2, which ensured that ^{me}G persisted in the DNA. Treatment with different doses of the methylating agent showed that replication was deleteriously affected only at MNU concentrations exceeding 80 mM (Fig. 3D); the observed protein recruitment was thus not the result of blocked replication. Similar experiments carried out with MNNG-treated chromatin (Fig. S3A and B) confirmed that, at doses that do not affect DNA replication, MMR proteins are recruited to methylated chromatin together with replication factors, and persist because of the presence of ^{me}G in the DNA.

In *in vitro* assays, MMR proteins were shown to directly interact with the checkpoint machinery upon ^{me}G recognition [17,18]. In contrast, our *in vivo* data indicated that activation of DNA damage signaling required replication. In order to gain more insight into this phenomenon, we examined the recruitment and activation of DNA damage signaling factors to MNU-treated sperm chromatin during a single round of replication in the naturally synchronous Patrin-2 pre-treated XEE. As control we used either untreated chromatin, or chromatin treated with MMS, which is known to elicit a potent checkpoint response in the extract [34]. As shown in Fig. 3E, treatment with 1% (110 mM) MMS resulted in the ready recruitment of DNA damage signaling machinery. In contrast, no comparable recruitment could be seen to chromatin treated with MNU, even at a dose of 80 mM, which led to MMR protein recruitment comparable to that seen with 1% MMS and which would have generated similar amount of methylation damage other than ^{me}G (predominantly N^7-methylguanine and N^3-methyladenine). At this dose, replication efficiency was beginning to decrease (Fig. 3D), while phosphorylation of Chk1, a downstream target of ATR became detectable and continued to increase up to 500 mM MNU (Fig. 3F and G), a concentration that severely inhibited replication (Fig. 3D). Moreover, Chk1 phosphorylation became Mgmt/Patrin-2 independent (Fig. 3G) and therefore independent of MMR-mediated processing of ^{me}G. Similar results were obtained with MNNG (Fig. S3C).

To test whether the presence of ^{me}G-containing gaps impairs replication in genomic DNA, we carried out a re-replication reaction with untreated- or MNU-treated sperm chromatin. The chromatin was first allowed to replicate in Patrin-2 pretreated extracts, isolated, and incubated with fresh extracts that were either pretreated with Patrin-2 or only with DMSO. As shown in Fig. 3H, re-replication of the ^{me}G-containing chromatin in the Patrin-2 pretreated extracts was severely impaired, but was unaffected in mock (DMSO) treated ones.

4. Conclusions

X. laevis egg extracts represent an *in vitro* experimental system that permits the dissection of biological processes that can otherwise be studied only *in vivo*. In this study, we set out to deploy this system to help us identify key processes and intermediates that bring about the cytotoxicity of ^{me}G residues generated by S_N1-type methylating agents. We demonstrate here that the presence of ^{me}G in DNA does not hinder replication, but that both ^{me}G/C and ^{me}G/T mispairs activate MMR. In contrast to a previous report [35], we show that, analogously to human cell extracts, the MMR process is nick-directed. We also show that ^{me}G in the template strand is not removed by MMR-mediated excision, but that its processing results in the generation of single-stranded gaps opposite the ^{me}G residues. The sluggish repair of these gaps is suggestive of iterative MMR-dependent resection/resynthesis. Unexpectedly, the checkpoint machinery activation was only weak, which suggested that the gaps were possibly shielded from signaling and thus unlikely to trigger a cell cycle arrest during the first replication round. However, the structures remaining in the DNA following one round of replication inhibited the subsequent replication round. This is consistent with the finding that the cell cycle arrest induced by S_N1-type methylating agents *in vivo* is

activated only after the second S phase, possibly through collision of the gaps with replication forks, which would give rise to double strand breaks that would activate the ATR-dependent G_2/M checkpoint and subsequently also apoptosis. Taken together, the above evidence provides further experimental support for the hypothesis that MMR-mediated processing of ^{me}G-containing mismatches arising during DNA replication rather than direct signaling of these mispairs to the apoptotic machinery is the underlying cause of the toxicity of S_N1-type methylating agents.

Conflict of interest

None declared.

Acknowledgements

The authors would like to express their gratitude to Odete Pereira Alves for able technical assistance, to Masahiro Akiyama, Julian Blow, Karlene Cimprich, William Dunphy, Howard Lindsay, Shou Waga and Juergen Thomale for the gift of antibodies, Stephanie Felscher for recombinant Exo1, Svenja Kaden for assistance with NPE preparations, Todor Anguelov for the modified oligonucleotides and Geoff Margison for Patrin-2. We also gratefully acknowledge the financial support of the Associazione Italiana per la Ricerca sul Cancro (AIRC) and the Giovanni-Armenise Foundation to V.C., and of the Swiss National Science Foundation (grant nos. 310030B-133123 and 31003A-149989) to J.J.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep.2015.01.014.

References

[1] J. Genschel, S.J. Littman, J.T. Drummond, P. Modrich, Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha, J. Biol. Chem. 273 (1998) 19895–19901.
[2] F. Palombo, I. Iaccarino, E. Nakajima, M. Ikejima, T. Shimada, J. Jiricny, hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA, Curr. Biol. 6 (1996) 1181–1184.
[3] N. Constantin, L. Dzantiev, F.A. Kadyrov, P. Modrich, Human mismatch repair: reconstitution of a nick-directed bidirectional reaction, J. Biol. Chem. 280 (2005) 39752–39761.
[4] W.H. Fang, P. Modrich, Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction, J. Biol. Chem. 268 (1993) 11838–11844.
[5] J. Jiricny, The multifaceted mismatch-repair system, Nat. Rev. Mol. Cell Biol. 7 (2006) 335–346.
[6] D. Fu, J.A. Calvo, L.D. Samson, Balancing repair and tolerance of DNA damage caused by alkylating agents, Nat. Rev. Cancer 12 (2012) 104–120.
[7] P. Karran, Mechanisms of tolerance to DNA damaging therapeutic drugs, Carcinogenesis 22 (2001) 1931–1937.
[8] B. Kaina, M. Christmann, S. Naumann, W.P. Roos, MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents, DNA Repair (Amst.) 6 (2007) 1079–1099.
[9] P. Karran, M.G. Marinus, Mismatch correction at O^6-methylguanine residues in *E. coli* DNA, Nature 296 (1982) 868–869.
[10] G. Aquilina, M. Crescenzi, M. Bignami, Mismatch repair, G(2)/M cell cycle arrest and lethality after DNA damage, Carcinogenesis 20 (1999) 2317–2326.
[11] M.T. Hawn, A. Umar, J.M. Carethers, G. Marra, T.A. Kunkel, C.R. Boland, M. Koi, Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint, Cancer Res. 55 (1995) 3721–3725.
[12] E. Papouli, P. Cejka, J. Jiricny, Dependence of the cytotoxicity of DNA-damaging agents on the mismatch repair status of human cells, Cancer Res. 64 (2004) 3391–3394.
[13] S.J. York, P. Modrich, Mismatch repair-dependent iterative excision at irreparable O^6-methylguanine lesions in human nuclear extracts, J. Biol. Chem. 281 (2006) 22674–22683.
[14] N. Mojas, M. Lopes, J. Jiricny, Mismatch repair-dependent processing of methylation damage gives rise to persistent single-stranded gaps in newly replicated DNA, Genes Dev. 21 (2007) 3342–3355.
[15] S. Quiros, W.P. Roos, B. Kaina, Processing of O^6-methylguanine into DNA double-strand breaks requires two rounds of replication whereas apoptosis is also induced in subsequent cell cycles, Cell Cycle 9 (2010) 168–178.
[16] L. Stojic, N. Mojas, P. Cejka, M. Di Pietro, S. Ferrari, G. Marra, J. Jiricny, Mismatch repair-dependent G2 checkpoint induced by low doses of SN1 type methylating agents requires the ATR kinase, Genes Dev. 18 (2004) 1331–1344.
[17] K. Yoshioka, Y. Yoshioka, P. Hsieh, ATR kinase activation mediated by MutSalpha and MutLalpha in response to cytotoxic O^6-methylguanine adducts, Mol. Cell 22 (2006) 501–510.
[18] Y. Liu, Y. Fang, H. Shao, L. Lindsey-Boltz, A. Sancar, P. Modrich, Interactions of human mismatch repair proteins MutSalpha and MutLalpha with proteins of the ATR-Chk1 pathway, J. Biol. Chem. 285 (2010) 5974–5982.
[19] N. Pabla, Z. Ma, M.A. McIlhatton, R. Fishel, Z. Dong, hMSH2 recruits ATR to DNA damage sites for activation during DNA damage-induced apoptosis, J. Biol. Chem. 286 (2011) 10411–10418.
[20] Y. Hashimoto, V. Costanzo, Studying DNA replication fork stability in Xenopus egg extract, Methods Mol. Biol. 745 (2011) 437–445.
[21] P.J. Lupardus, C. Van, K.A. Cimprich, Analyzing the ATR-mediated checkpoint using Xenopus egg extracts, Methods 41 (2007) 222–231.
[22] J.P. Chong, P. Thommes, A. Rowles, H.M. Mahbubani, J.J. Blow, Characterization of the Xenopus replication licensing system, Methods Enzymol. 283 (1997) 549–564.
[23] Y. Kubota, H. Takisawa, Determination of initiation of DNA replication before and after nuclear formation in Xenopus egg cell free extracts, J. Cell Biol. 123 (1993) 1321–1331.
[24] J. Walter, L. Sun, J. Newport, Regulated chromosomal DNA replication in the absence of a nucleus, Mol. Cell 1 (1998) 519–529.
[25] J.J. Blow, R.A. Laskey, Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs, Cell 47 (1986) 577–587.
[26] K. Taira, S. Nakamura, K. Nakano, D. Maehara, K. Okamoto, S. Arimoto, D. Loakes, L. Worth Jr., R.M. Schaaper, K. Seio, M. Sekine, K. Negishi, T. Negishi, Binding of MutS protein to oligonucleotides containing a methylated or an ethylated guanine residue, and correlation with mutation frequency, Mutat. Res. 640 (2008) 107–112.
[27] P. Cejka, N. Mojas, L. Gillet, P. Schar, J. Jiricny, Homologous recombination rescues mismatch-repair-dependent cytotoxicity of S(N)1-type methylating agents in *S. cerevisiae*, Curr. Biol. 15 (2005) 1395–1400.
[28] D.R. Duckett, J.T. Drummond, A.I. Murchie, J.T. Reardon, A. Sancar, D.M. Lilley, P. Modrich, Human MutSalpha recognizes damaged DNA base pairs containing O^6-methylguanine, O^4-methylthymine, or the cisplatin-d(GpG) adduct, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 6443–6447.
[29] A.M. Schmitt, C.D. Crawley, S. Kang, D.R. Raleigh, X. Yu, J.S. Wahlstrom, D.J. Voce, T.E. Darga, R.R. Weichselbaum, B. Yamini, p50 (NF-kappaB1) is an effector protein in the cytotoxic response to DNA methylation damage, Mol. Cell 44 (2011) 785–796.
[30] L.A. Lindsey-Boltz, M.G. Kemp, J.T. Reardon, V. DeRocco, R.R. Iyer, P. Modrich, A. Sancar, Coupling of human DNA excision repair and the DNA damage checkpoint in a defined in vitro system, J. Biol. Chem. 289 (2014) 5074–5082.
[31] C.A. MacDougall, T.S. Byun, C. Van, M.C. Yee, K.A. Cimprich, The structural determinants of checkpoint activation, Genes Dev. 21 (2007) 898–903.
[32] H. Hombauer, C.S. Campbell, C.E. Smith, A. Desai, R.D. Kolodner, Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates, Cell 147 (2011) 1040–1053.
[33] H.E. Kleczkowska, G. Marra, T. Lettieri, J. Jiricny, hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci, Genes Dev. 15 (2001) 724–736.
[34] M.P. Stokes, R. Van Hatten, H.D. Lindsay, W.M. Michael, DNA replication is required for the checkpoint response to damaged DNA in Xenopus egg extracts, J. Cell Biol. 158 (2002) 863–872.
[35] I. Varlet, M. Radman, P. Brooks, DNA mismatch repair in Xenopus egg extracts: repair efficiency and DNA repair synthesis for all single base-pair mismatches, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 7883–7887.

Supplementary Information

DNA substrates and *X. laevis* MMR assay

The substrates were generated as described previously (Baerenfaller et al, 2006). Briefly, the hetero- and homoduplexes were constructed by primer extension using the oligonucleotides listed below as primers and the single-stranded phagemid DNA as template. pRichi2850top*Sal*I (pRichi2850t *Sal*I) pRichi2850bottom *AclI* (pRichi2850b) and pRichi2850top *AclI* (pRichi2850t) that create 5' top-nicked or 3' bottom-nicked substrates were used as the ssDNA templates. A single mutation abrogating an *Acl*I site and creating a *Sal*I site was used to generate the pRichi2850top *Sal*I phagemid. After primer extension and ligation, the desired supercoiled heteroduplex substrates were purified on CsCl gradients.

Mismatch-provoked excision/resynthesis assays were performed as previously described (Varlet et al, 1990) in the presence of 100 ng DNA substrate and 26 µl S phase extract in a total volume of 30 µl in a buffer containing 15 mM Hepes-KOH pH 7.4, 15 mM potassium glutamate, 8.75 mM magnesium acetate, and 2 µCi of [α-^{32}P]dCTP. The extracts were incubated at 23°C for 45 min. The reaction was stopped by adding 70 µl STOP solution containing 1 mM EDTA, 3% SDS and 5 mg/ml proteinase K (Roche). The samples were incubated at 55°C for 30 min, purified on MinElute columns (Qiagen) and subjected to restriction digests in the presence of RNase (5prime) at 37°C overnight. The digested DNA was cleaned up again and resolved on 1% agarose/TAE gels stained with GelRed.

Primers

All primers were obtained from Microsynth (Balgach, Switzerland). The sequences are indicated below. The *Sal*I restriction site (GTCGAC) is underlined. The mispaired residue is highlighted in bold.

C/G-*Sal*I primer:

5' CCAGACGTCTGT**C**GACGTTGGGAAGCTTGAG 3'

T/G-*Sal*I primer:

5' CCAGACGTCTGT**T**GACGTTGGGAAGCTTGAG 3'

^{me}G/C and ^{me}G/T *-SalI primer:

5' CCAGACGTCTGTC^{me}**G**ACGTTGGGAAGCTTGAG 3'

* pRichi2850top *AclI* was used for this substrate to pair a T with ^{me}G

Production of recombinant proteins

Fragments of *X. leavis* MMR proteins and full-length *X. laevis* Mgmt were PCR-amplified from EST clones obtained from Imagenes using primers containing restriction sites for *Nco*I and *Xho*I. The PCR products were then cloned in a pET28B(+) vector (Novagen). 6xHis-tagged protein fragments (Msh2F1: 258-404 aa, Mlh1F1: 615-744aa, Msh6F1: 128-249aa, Exo1F1: 344-450aa) were expressed in BL21 cells (Invitrogen) and purified on Ni-NTA agarose columns (Qiagen) according to the manufacturer's protocol.

Antibodies

Rabbit anti-*X. laevis* Msh2, Msh6, Mlh1 and Exo1 antisera were generated by the Institute of Laboratory Animal Science (University of Zurich) against recombinant 6xHis-tagged protein fragments described above. Two rabbits per fragment were injected. Immunoglobulins specific for the *X. laevis* MMR proteins were affinity-purified using Hitrap NHS-activated HP (GE healthcare) according to standard protocols. Only the Msh6 and Exo1 sera were suitable for immunodepletions.

Rabbit polyclonal antibodies against *X. laevis* TopBP1, Atr, Rad9, Rad1, Hus1, Rad17, Orc2, Pol δ and Pol ε were gifts from Masahiro Akiyama (Nara Institute of Science and Technology), Julian Blow (Dundee University), Karlene Cimprich (Stanford University), William Dunphy (California Institute of Technology), Howard Lindsay (Lancaster University) and Shou Waga (Japan Women's University). Antibodies against human PCNA (Santa Cruz), MCM7 (Sigma), CHK1 (Santa Cruz), polymerase α (Abcam) and RPA32 subunit (Abcam), which cross-react with the *X. laevis* proteins, were obtained commercially. The antibody that recognizes phospho-S_{345} of *X. laevis* p-Chk1

was purchased from Cell Signaling Technology. Anti-^{me}G (EM2-3) antibody was a kind gift of Jürgen Thomale (Duisburg-Essen University).
Western blot procedures were performed using ECL or ECL advanced (GE Healthcare). Antibodies for Western blots were used at 1:1000 dilutions, unless otherwise specified.

Dot blots

The dot blot procedure was described previously in "Technologies for detection of DNA damage and mutations" (G.P. Pfeifer, 1996 edition). Briefly, 200 ng of pRichi2850t *Sal*I substrates were heat-denatured at 100°C for 10 min, 2 M ammonium acetate was added to a final concentration of 1 M and the samples were blotted onto pre-wetted nitrocellulose. The membrane was rinsed in 5 x SSC (0.75 M sodium chloride and 75 mM trisodium citrate) and blocked for 1 h in PBS/0.5% Casein/0.1% deoxycholate for 1 h at RT. Anti-^{me}G antibody (EM2-3) was used at 1:10 000 dilution for 1 h at RT in the same solution. Membranes were washed with PBS/160 mM NaCl/0.1% Triton X-100 for 3 x 5 min, incubated with a secondary anti-mouse antibody, washed further 3 times and developed.

Band-shift assays

The oligonucleotide heteroduplexes were created by annealing the 5' labeled oligonucleotide ^{me}G, primer for synthesis of the top (complementary) strand of the heteroduplex, with its complementary oligonucleotides containing a C or a T opposite the ^{me}G. Where indicated, the annealed substrates were pre-treated with Mgmt. The binding reaction mixtures (50 mM Hepes pH 7.5, 0.5 mM EDTA, 10% glycerol, 0.5 mM DTT, 0.5 mg/ml BSA and 20 ng poly(dI-C) contained 40 fmol duplex ^{me}G/C or ^{me}G/T and human MSH2/MSH6 (250 ng). Protein-bound substrates were separated from the free probes by electrophoresis on 6% native polyacrylamide gels eluted with TAE.

***Xenopus laevis* MSH6 immunodepletion**

To prepare Msh6-depleted extracts, 100 μl of the extract were incubated 3 times for 30 min with 33 μl protein A-Sepharose beads 4B fast flow (Sigma)

coupled to 100 μl of Msh6 antiserum. Mock depletion was performed using beads coupled to the pre-immune serum. Coupling to the beads was performed overnight at 4°C without dilution. After the coupling reaction, the beads were washed extensively with EB buffer. The extract was reconstituted with 1 μg hMutSα where indicated.

Supplementary Figure Legends

Figure S1. Purification of recombinant *X. laevis* Mgmt, quality control of ^{me}G-containing substrates in *XEE,* Mgmt inhibitor testing and antibody generation. A, Sequences of pRichi2850 substrates depicting the position of the T/G, ^{me}G/C and ^{me}G/T mismatches with respect to the restriction sites used in this study (GTCGAC, *Sal*I, TTCGAA, *Hin*dIII). **B**, SDS-PAGE of full-length 6xHis-tagged Mgmt (25kDa, 1 μg/μl). **C**, Band shift assay with hMutSα and ^{me}G/C or ^{me}G/T substrates. Treatment of the substrates with Mgmt prior to incubation with hMutSα converted the oligos to C/G and T/G, respectively. **D**, Efficiency of the Mgmt inhibitors O^6-benzylguanine (BG) and O^6-(4-bromothenyl)guanine (Patrin-2) in XEE. HSS was incubated for 10 min with increasing concentrations of either BG or Patrin-2 and then the ^{me}G/C substrate was added to the mix for a further 30 min. This was followed by *Sal*I cleavage of the substrate and analysis on 0.8% agarose/TAE gels to determine the most effective dose of the inhibitor. Lack of cleavage indicates presence of ^{me}G/C in the *Sal*I site. Left panel: GelRed-stained gel of a representative assay for BG efficiency. Right: Quantification of experiments done with BG and Patrin-2. **E**, BLAST search showing degree of conservation between *H. sapiens* and *X. laevis* MMR proteins of interest. **F**, Coomasie-stained 6xHis-tagged MMR protein fragments resolved on a 15% SDS-PAGE. 3 μg of purified fragment F1 was loaded per lane (Msh2F1: 1.5 μg/μl, Msh6F1: 1.5 μg/μl, Mlh1F1: 3 μg/μl, Exo1F1: 1.2 μg/μl). M, size marker in kDa. **G**, Western blot analysis of the *X. laevis* egg extract (NPE, 1μl) with the affinity-purified MMR antibodies using the following dilutions (Msh2: 1:5000, Msh6: 1:5000, Mlh1: 1:500, Exo1: 1:1000).

Figure S2. DNA damage signaling, gapped substrate preparation and restriction digest analysis.

A, Kinetic analysis of checkpoint activation (p-Chk1) during replication of the indicated substrates in HSS+NPE. Aliquots removed at the indicated times were resolved on a 7.5% SDS-PAGE and probed for p-Chk1 (S345) and total Chk1. Time point 0 indicates addition of NPE to the HSS/DNA mix. **B**, GelRed-stained 0.8% agarose/TAE gel showing the following samples (from left to right): 1kb DNA ladder, single-stranded phagemid pRichi350 (200 ng), supercoiled pRichi350 (300 ng), bottom-nicked pRichi350 (150 ng) and gapped pRichi350 substrate (150 ng). The latter was made by incubating 10 μg of the bottom-nicked ^{me}G/C substrate with recombinant human Exo1 (28.5 nM) for 5 min at 37˚C. When nicked by Nt.*Bst*NB1, the nick occurs at position 350 in the 3197 bp pRichi350 substrate (308 bp from ^{me}G, not depicted). **C**, Top panel: Schematic representation of the restriction sites for *Not*I (30 bp from ^{me}G), *Nde*I (730 bp from ^{me}G) and *Xmn*I (1,3 kb from ^{me}G) in the pRichi350 substrate with regard to the ^{me}G (triangle). Bottom panel: GelRed-stained 0.8% agarose/TAE gel showing restriction digest analysis of the gapped pRichi350 substrate (150 ng). The restriction enzymes used are indicated above the gel. The result shows that the gaps extended past the *Not*I site but did not reach the *Nde*I site.

Figure S3. Recruitment of MMR and replication factors to chromatin and DNA damage signaling induced by MNNG. A, Kinetic analysis of replication of undamaged- and 1.5 mM MNNG-damaged sperm chromatin in the absence or presence of the Mgmt inhibitor Patrin-2. Recruitment of Msh2, MCM7 and pol-ε to chromatin was monitored by Western blotting. **B**, ^{me}G-dependent recruitment of MMR proteins to chromatin. Binding of mismatch repair proteins Msh2, Msh6 and Mlh1 to undamaged- or MNNG-damaged chromatin in the absence or presence of the Mgmt inhibitor Patrin-2. The samples were assayed after 120 min. E, extract. **C**, Checkpoint activation in XEE in response to the indicated MNNG concentrations. Undamaged- or MNNG-damaged sperm chromatin was replicated in S phase extracts in the

absence or presence of Patrin-2. Nuclear proteins were isolated after 120 min and probed for p-Chk1 (S345) and Chk1.

References

Baerenfaller K, Fischer F, Jiricny J (2006) Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro. *Methods Enzymol* **408:** 285-303

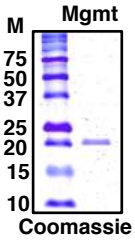
Varlet I, Radman M, Brooks P (1990) DNA mismatch repair in Xenopus egg extracts: repair efficiency and DNA repair synthesis for all single base-pair mismatches. *Proc Natl Acad Sci U S A* **87:** 7883-7887

A

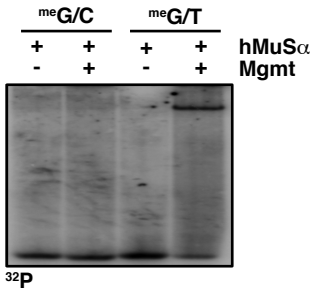
pRichi2850 substrate	Position of mismatch
C/G:	GTCGACG TTCGAA CAGCTGC AAGCTT
T/G:	GTTGACG TTCGAA CAGCTGC AAGCTT
^{me}G/C:	GTCXACG TTCGAA CAGCTG CAAGCTT
^{me}G/T:	GTCXAC GTTCGAA CAGTTGCAA GCTT

X: O^6-methylguanine (^{me}G)

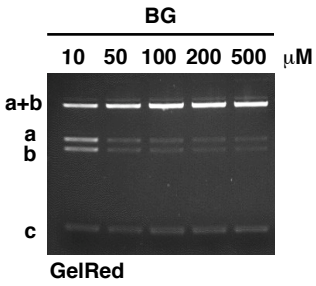
B



C



D



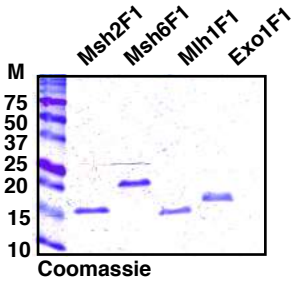
Enzyme cleavage (%) at ^{me}G/C

μM	BG	Patrin-2
10	46.4	23.9
50	16.6	9.6
100	10.5	8.1
200	7.9	7.7
500	7.7	8.2

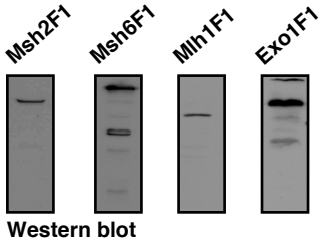
E

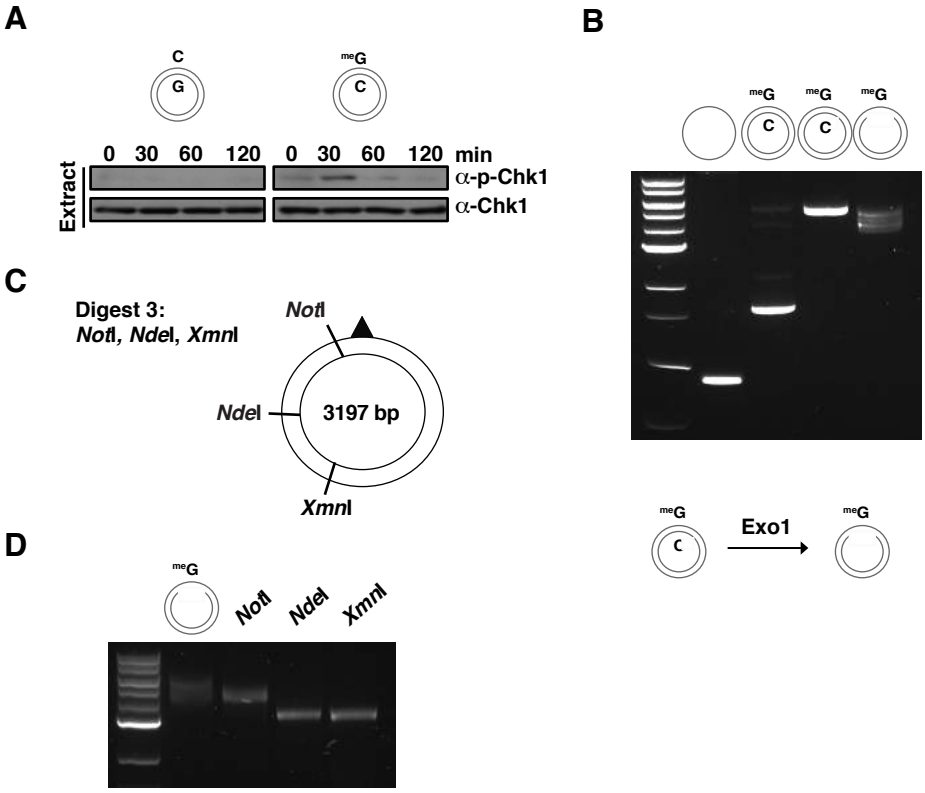
Xenopus MMR protein	Identity to human (%)
Msh2	79
Msh6	66
Mlh1	86
Exo1	52

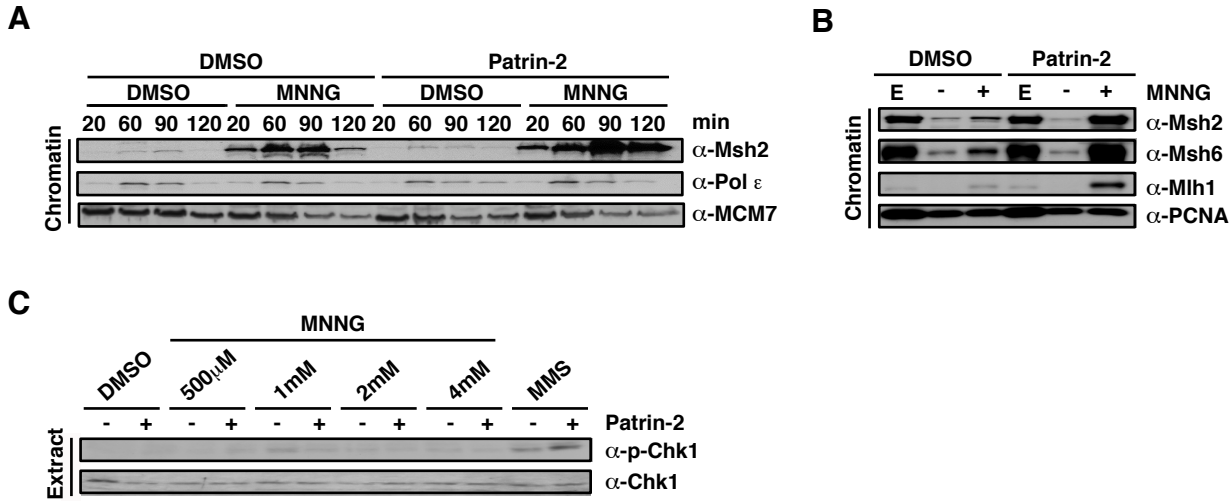
F



G




A
C
G
meG
C
0 30 60 120 0 30 60 120 min
Extract
α-p-Chk1
α-Chk1
B
meG meG meG
C
C
meG
C
Exo1
meG
C
Digest 3:
NotI, NdeI, XmnI
NotI
NdeI
3197 bp
XmnI
D
meG
NotI
NdeI
XmnI


A
DMSO
Patrin-2
DMSO
MNNG
DMSO
MNNG
20 60 90 120 20 60 90 120 20 60 90 120 20 60 90 120
min
Chromatin
α-Msh2
α-Pol ε
α-MCM7
B
DMSO
Patrin-2
E - + E - +
MNNG
Chromatin
α-Msh2
α-Msh6
α-Mlh1
α-PCNA
C
MNNG
DMSO
500μM
1mM
2mM
4mM
MMS
- + - + - + - + - + - +
Patrin-2
Extract
α-p-Chk1
α-Chk1